

Europe should welcome diversity

With the arrival of three new members to the European Union, universities could be the centrepiece of a common purpose providing Europe with cultural cement.

THE best thing to have happened to the European Union since the ratification of the Maastricht Treaty last November is the prospect that Austria, Finland and Sweden will now become members. The three governments agreed terms with Brussels some weeks ago; later this year they will each hold a referendum to win assent to what is planned. If all goes well, the result will be not merely that the European Union spans the whole of Europe in two orthogonal directions, from North to South and East to West, but that the accession of these particular new members will lend the European Union a welcome degree of further diversity. Austria is not a little Germany to the East, but has become an enterprising and innovative economy in an historic setting. The two Scandinavian countries will similarly bring to the rest of Europe substantial and successful research enterprises and, more important, strong traditions of rationality in the management of public affairs that are not always conspicuous elsewhere.

But why should Europe welcome diversity when an excess of that commodity is one of its perennial problems? How will the European Commission ever be able to create a sense of European "identity" (one of its goals) when new members, speaking languages not yet on the translation roster at Brussels, are forever arriving on the doorstep? The answer is simple; Europe is not a "nation" (as are, for example, the Scots and the Sioux), and cannot become one even in the next millennium. Nor is there an easy way of smoothing over the differences of language within Europe, either by pretending that they do not matter or by fostering polyglotism.

The European Commission's line for the past few years has been to identify and then proclaim common elements in European cultural history. But even that is not straightforward. Knowing that Roman armies once occupied almost everywhere in Europe does not create a common sense of purpose among the descendants of the survivors of those regimes. Indeed, as social anthropologist Chris Shore from the University of London has recently argued (*Man* 28, 779–800; 1993), preoccupation with common culture may be as powerful a spur to xenophobia as to identity. A related difficulty is that the past, important though it is, by necessity gives people preoccupied with jobs and other contemporary problems a sense of being separated from a more delectable past.

So where should the European Commission be looking for cultural cement? For one thing, in the present. For

another, in enterprises that promise something for now and for the future. So why not make the cause of the European university the centrepiece of an attempt to give all Europeans a sense of common, in multilingual, purpose? In some respects, the commission has had this idea already. There is an excellent and substantial programme for the transfer of students among European universities, for example. Many of the commission's research programmes are similarly angled at multinationality. But there has not yet been a serious attempt to harness these sporadic initiatives into a coherent whole. The arrival on the scene of three new and interesting member states should be a spur to that end.

Several issues need attention, and can be given it now that the Maastricht Treaty gives the European Union competence in the field. The temptation will be to devise ways in which students can become even more mobile than they are at present, perhaps by legislating for modular courses that students can follow here or there. But that is the wrong direction; universities also need diversity. But mobility for teachers, only partly addressed by the commission's present programme on human mobility, is a more urgent need. So are the tools of scholarship, from libraries in the old-fashioned sense to information super-highways in the modern. But while initiatives such as these could make academics feel good, would they not alienate the less fortunate who worry about jobs for themselves and their children. When the participation rate of the young in Europe is around 30 per cent, that would be the opposite of the truth. □

FDA lights Smoking Fire

The US Food and Drug Administration is thinking about defining cigarettes as drug "delivery devices".

IN what is a bold move by any standards, David Kessler, the commissioner of the US Food and Drug Administration (FDA) has ventured the opinion that cigarettes might be legally defined as drug delivery devices, thereby subject to FDA regulation. Inasmuch as cigarettes deliver an addictive drug—namely nicotine—FDA regulation would be tantamount to a ban on the sale of cigarettes nationwide.

In a controversial letter to an antismoking coalition, Kessler recently said that there is evidence manufacturers "commonly add nicotine to cigarettes." This is so. After all, people do not smoke tobacco leaves whole. Rather, manu-



facturers separate leaves and stems, extract the nicotine, pulverise the plant products and flatten them into sheets that are rolled into a cigarette. The extracted nicotine is sprayed onto the sheets of tobacco.

It is Dr Kessler's contention that some manufacturers deliberately increase the nicotine concentration to increase the likelihood that smokers will become addicted. The FDA's antismoking gambit, which Dr Kessler hopes will be debated at congressional hearings, comes at a time when the United States is well on its way to becoming a nearly smoke-free society.

Smoking is prohibited on most commercial airlines, including those flying coast-to-coast. More and more office buildings are restricting smoking. As a result, die-hard smokers are forced outside where they can be seen in rain and cold, huddling by the doorway.

It is now common that formal restaurants to have no-smoking sections. And this month, McDonald's (home to the ubiquitous Big Mac) banned smoking, while the National Council of Chain Restaurants declared its support for pending legislation, (introduced by Representative Henry Waxman (Democrat-California), to prohibit smoking in virtually all public places.

Much is being written these days about the tragedy of young people taking up smoking as older generations quit or die. A public ban would certainly send a strong message.

This antismoking fervour, which originates from the correct presumption that smoking is generally bad for smokers' health, is now ignited by the less conclusively proven hypothesis that so-called secondary or environmental smoke is a serious health hazard. For instance, a recent literature survey in *JAMA* (271, 612-617) reported an association between negative findings and research support from the tobacco industry whereas independent reports tend to confirm a link between environmental smoke and respiratory, cardiovascular and other ailments in nonsmokers. However, this does not confirm the scientific rigor of the data. Meanwhile, challenges to the Environmental Protection Agency's conclusions about the extent of danger from environmental smoke have yet to be resolved.

Nevertheless, it is hard to argue in favour of tobacco addiction. Unlike alcohol (society's other sanctioned yet often addictive drug), which may provide some protection to the cardiovascular system and is relaxing in moderation, there is no case to be made that smoking confers any benefits.

By even suggesting that cigarettes could be regulated by the FDA as an addictive drug, Commissioner Kessler has raised the ante in the long-standing impasse between the medical establishment and the powerful, wealthy tobacco industry whose contributions to congressional campaigns have earned it the loyalty of members of both the House and Senate. Certainly one of the US government's great contradictions is that it adjures against smoking on the one hand while subsidizing tobacco growers on the other. If the FDA can force a showdown on this issue, so much the better. □

Animals in research

Tufts vets seek a middle ground between researchers and those who oppose using animals in science.

THE numbers of animals used in research, both in the United States and Europe, has declined significantly during the past decade — a fact that should hearten those who are concerned about the welfare of animals. According to "the animal research controversy," a comprehensive report from the Center for Animals and Public Policy at the Tufts University School of Veterinary Medicine in Massachusetts, the numbers have been falling since the 1970s, particularly in Great Britain and Europe where good data indicate a decline in the range of 20 per cent to 50 per cent. British researchers, for example, used some 3 million animals in 1992, down from 6 million in 1975. Data on the numbers of animals used in the United States are less accurate, but a good guess is an overall decrease of about 25 per cent since 1985.

But statistics alone will do little to quell the acrimonious debate about the morality (or even necessity) of using animals as surrogates in the study of human biology. Examples of diseases that can be treated because of animal studies are valid but not convincing to activists. The debate really hinges on issues such as the relative value to human beings of inflicting pain and distress on animals in order to gain data, or personal views about the cognitive lives of dogs, cats, and other species used in science. "Laboratory animal research causes less pain and distress than implied by animal protection literature but more animal pain and distress than claimed by research advocates," said Franklin Loew, dean of Tufts and a coauthor of the study. Loew also aptly noted that people on either side of the issue tend to mischaracterize their opponents ("scientists as sadists," for instance, or animal rights activists as "emotional fanatics").

Is there a solution? First, it can be said that there is nothing substantially new in the debate that has been going on for a century. It is true that it is now possible to obtain certain kinds of information (toxicity data, for instance) from alternative test systems like bacteria. But that does not go to the heart of the matter which is the value of animal life over the lives and health of human beings.

The Tufts group has hopefully called for the creation of an officially sanctioned forum, where both sides could come together. But in an arena characterized by *ad hominem* verbal attacks and even criminal violence, it is hard to imagine negotiated conflict resolution around a table in Washington or London. The most adamant critics and defenders of animal research show no signs of wanting to compromise. Nevertheless, there are indications that some groups are willing to talk. Loew has done a service by acknowledging that animal experimentation can be painful. In return, representatives of the Humane Society of the United States have called the idea of a forum "a good thing". That in itself is progress. □

British research council seeks to cut role of collective peer review

London. A radical change in the way that research grant applications are evaluated for funding, eliminating the role of peer review committees and increasing the discretionary powers of programme managers, is being proposed for Britain's new Engineering and Physical Sciences Research Council (EPSRC).

The move is part of a government-inspired strategy both to streamline the operation of the research councils, and to increase the weight given in funding decisions to the perception of a research project's potential contribution to wealth creation.

At present, all grant applications submitted to the Science and Engineering Research Council (SERC) are assessed by the members of one of the SERC's specialist review committees. The committee members collectively assign a score to each application, and those receiving the highest scores are put forward for funding.

A radical change is now being proposed for the EPSRC, which will take over a large proportion of the responsibilities of the SERC from the beginning of next month, in line with the recommendations of last year's white paper on science. (Most of the SERC's

remaining activities will be taken on by the new Particle Physics and Astronomy Research Council.)

The proposed changes, which have been drawn up by SERC officials, will be put to the new council next month. These suggest that the review committees should be abolished, and that future applications for research grants will be assessed by programme managers, drawing on advice from outside referees.

"Peer review in the way that the SERC used to run it is being changed," a spokesman for the council said on Monday. "We are moving to a managed system, where peer reviewers will still be used as advisers, but will not have the delegated executive authority [to make judgements on applications] that they have at present."

The changes are part of a general shake-up of the way that research councils operate which is being demanded by the Office of Science and Technology, and steered through by John Cadogan, the new director-general of research councils.

Last month, for example, William Waldegrave, the cabinet minister responsible for science, announced that half of an

extra £15 million which is to be spent on new research programmes designed to further the white paper's objectives is to be found from efficiency savings within the research councils.

The prospect of cutting down on the bureaucracy which currently accompanies the grant review process is likely to be welcomed in the scientific community, many of whose members complain of the amount of time and paperwork they have to spend on current procedures for assessing applications.

At the same time, however, there is concern that the new procedures could undermine the traditional peer review process, and reduce the weight given to scientific quality in comparison to other criteria for assessing research proposals — in particular their explicit industrial relevance.

"Many academic scientists will be unhappy with this change," one SERC committee member said last week. "Despite the amount of work involved, we like the existing system because we feel it does a reasonable job, and doubt whether a different way of doing things will be an improvement."

Some of the supporters of the change point out that giving greater responsibilities to programme managers will bring the EPSRC's system for allocating grants closer to that of other countries, for example the US National Science Foundation (NSF).

But its critics point to the significant difference that, in the United States, most scientists are faced with a range of potential sources of federal support. In contrast, the SERC is the sole source of public funding (apart from that provided through the university funding councils) for university scientists in many fields.

In addition, NSF programme managers tend to be academic scientists working on secondment to the agency. In contrast, the programme managers which the SERC is proposing should now be responsible for handling grant applications will be permanent SERC employees.

"For such a system to work, the programme managers must be individuals who can command respect in the scientific community," says John Mulvey, secretary of the pressure group Save British Science.

SERC officials insist that the new procedures are an integral part of the 'culture change' that will be involved in the council's transformation into its new form, emphasizing the priorities outlined in the white paper and the council's responsibility to meet its new 'missions' of improving wealth creation and the quality of life.

But the proposed changes seem ►

Deep-sea probe draws up a blank screen

Tokyo. Japan's new unmanned deep-sea probe *Kaiko* has run into more bad luck. Last week, *Kaiko* set out to break the world depth record in the Challenger Deep off Guam. But after it had travelled down to about 10,900 metres, the television pictures it had been transmitting suddenly went blank 2 metres short of the sea floor, and the probe had to be hauled back to its mother ship, *Yokosuka*.

Kaiko will have to return to Japan for repairs, and a second attempt at breaking the 34-year-old record of 10,912 metres set by the Bathyscaphe *Trieste* in Challenger Deep in 1960 will have to wait until at least December. Last August, the hull of *Kaiko* was badly damaged during trials when it slammed into the side of the *Yokosuka* as it was being lowered into rough seas. The test in the Challenger Deep had to be delayed about seven months as a result of this accident.

The precise reason for last week's failure is unknown, but is believed to have been caused by a fault in the data transmission cable between the probe and the *Yokosuka*. The *Kaiko* has a manipulator arm that can be remotely operated to pick up samples spotted by the television cameras, and will therefore be able to perform

IMAGE
UNAVAILABLE FOR
COPYRIGHT
REASONS

Kaiko being prepared for dive.

some of the functions of a manned submersible, but at much greater depth. So far, ¥5,400 million (\$52 million) has been invested in its development by the Japan Marine Science and Technology Centre. □

Kyodo Photo Service

destined to generate considerable controversy in the scientific community, coinciding as they do with other indications that a concern for industrially relevant research may overshadow arguments about the need to devise a system for ensuring the maintenance of high scientific standards.

"It is a very dangerous step to take," says Lewis Moonie, the Labour Party's science spokesman. "The peer review system works fairly well, and something that works well should not be changed, while the purpose of the research councils is to fund basic research, not to plug the gaps caused by cut-backs elsewhere in the government's funding of applied research." **David Dickson**

● Richard Brook, Cookson Professor of Materials Science and head of the department of material at Oxford University, has been appointed the first chief executive officer of the new Engineering and Physical Sciences Research Council.

Brook is a specialist in ceramics and their application to mechanical engineering, and is currently chairman of the SERC's materials commission, as well as vice-president of the Institute of Materials.

His appointment completes the government's team of new research councils' heads, each now having an industrialist as a part-time chairman, and an academic scientist as a full-time chief executive officer.

Cancer geneticists sign deal with genome company

Washington. A collaborative agreement has been signed by a research team led by Bert Vogelstein and Kenneth Kinzler at Johns Hopkins University, Maryland, and the gene sequencing company Human Genome Sciences (HGS).

The deal will give HGS exclusive licensing rights to the outcome of research carried out at Johns Hopkins into the connection between colon cancer and genes sequenced at both HGS and its non-profit associate, Craig Venter's Institute for Genomic Research.

William Haseltine, the chairman of HGS, says the deal is no different from the agreement which any laboratory would use to protect its intellectual property. But he adds that it is the first of many such agreements which HGS hopes to sign.

Vogelstein — the top-cited scientist of 1993, primarily for his work on the genetic basis of cancer, according to statistics compiled by the Philadelphia-based newsletter *ScienceWatch* in which Kinzler was second — is reluctant to give further details about the deal. He says that its significance will become clearer when research connected with it, which is already completed and has been submitted to a US journal, is published.

C. M.

Fiscal limits cut impact of Clinton's technology dream

Washington. Al Gore, the US vice-president, last week launched a joint research and development initiative between the federal government and the US semiconductor industry which, he said, would be "of an entirely different order of magnitude" to similar activities already underway.

But the details of the initiative, which will involve the departments of defence, energy and commerce, suggest that it is likely to have a relatively small impact compared to that of the \$4.5 billion which the US semiconductor industry already spends each year on research.

Apart from creating yet another Washington oversight board — the Semiconductor Technology Council — the initiative will have three substantive elements, one old and two new.

The largest element will come from the Department of Defense, which will continue to invest about \$90 million a year in Sematech. This is a collaboration between government and industry which, despite much criticism, has been running since 1987, and has lately focused on helping US companies build the machines to make microchips.

This effort will now be complemented by the Department of Energy (DOE), which will spend around \$50 million over the next five years on a centre to help the microchip industry with computer simulation. The Department of Commerce will provide up to £25 million over this period through the National Institute of Standards and Technology (NIST) for a metrology centre to help the industry measure tiny distances. In each case, industry is expected to match the government's investment.

The new initiative has been generally welcomed by the electronics industry. But, as the US government slowly begins to steer its \$72 billion research budget away from fundamental science towards applied research, the key question remains whether initiatives of this type will actually produce significant economic benefits.

Last week, top managers from the industry who had gathered in Washington to help launch the new initiative praised Sematech's track record, and dutifully pledged that the advantages promised by Gore from the additional efforts will be forthcoming.

Craig Barrett, chief executive of Intel and president of the Semiconductor Indus-



Gibbons: need for 'paradigm shift'.

try Association, said Sematech had helped his industry "to pull itself together" in the face of the Japanese threat in the late 1980s. He added that the new deal would prove "as vital an influence as Sematech was over the last five years."

But few view Sematech as pivotal in the turnaround achieved by the US industry, which last year overtook Japan as market leaders for the first time since 1985. Sematech itself encountered problems early in its life, and some members pulled out as its original goal of getting rival firms to co-operate on chip development proved unrealistic.

The collaboration has since become more narrowly defined, concentrating on semiconductor manufacturing equipment, and is widely credited in the industry with improving the quality of such equipment built in the United States.

But if the administration really wanted to expand the Sematech model into a substantial inter-agency partnership, fiscal constraints seem to have prevented it from doing so. In practice, the amount of new money is trivial compared with the overall research and development budgets of the industry.

Despite funding restrictions, administration officials claim the new initiative will still have a significant impact. "We want to strengthen a partnership which has already become profound," Jack Gibbons, the president's science advisor, said last week.

Gibbons says that the semiconductor industry "is facing challenges which may call for a paradigm shift". As chips get smaller, he adds, "engineers will have to know quantum mechanics that were previously only of interest to physicists".

Energy secretary Hazel O'Leary said that the initiative differed from previous policy both in its multi-agency nature, and in the spirit in which it was entered into. "In the past five years it was a reluctant partnership," she said. "We are attesting to the fact that we like this partnership, and we need others like it." Although O'Leary talked of selecting a site for the DoE computer simulation centre, officials said the money was likely to be divided between the department's Sandia, Los Alamos and Lawrence Livermore laboratories.

The administration's hope is that the centres to be set up by NIST and the DoE will successfully focus on small niches which the industry might otherwise neglect.

Furthermore, the semiconductor industry may be feeling that, however hollow the idea of a 'research partnership' with government, a public expression of gratitude is a small price to pay for favours it has received elsewhere, for example in trade negotiations with the Japanese.

Colin Macilwain

Satellite delay adds to Italy's space blues

Munich. The launch of the high-resolution X-ray telescope XMM, one of the four 'cornerstone' missions of the European Space Agency's science programme, is likely to be delayed by at least a year if ESA adopts a proposal to ask for new bids for the project's main management contract.

This follows criticism by ESA officials of a tender for the ECU280 million contract to design and build the XMM — originally planned for launch in 1998 or 1999 — which had been submitted by a consortium of European companies headed by Italy's main space contractor Alenia Spazio.

The ESA executive has agreed that the tender is technically acceptable, but deemed it unacceptable on grounds of both cost and the proposed programme management. As a result, it has recommended to the ESA industrial policy committee, which meets next week (14 March), that the whole programme be put out again for tender.

Alenia is refusing to make any public comment on the affair. But the decision is a new blow to Italy's space community, whose role in ESA has already been weakened by the country's general political and economic difficulties, as well as by charges of corruption against the Italian Space Agency (ASI).

Italy seems to have promised in recent years a higher level of participation in ESA programmes than the country now finds it can afford. Such overcommitments, combined with last year's devaluation of the lire, mean that Italy is already having difficulty in paying its membership fees to ESA, which this year come to ECU370 million.

After discussions earlier this year with Umberto Colombo, the minister for research and universities, ESA has agreed to consider taking out a loan of up to US\$100 million to cover the expected deficits in Italy's contributions to the agency over the

next few years. This procedure, although unusual, is not unprecedented, and Colombo has promised that the loan would be paid back in 1996 and 1997, when he expects Italy's beleaguered space programme to be back on the rails.

Meanwhile, Colombo has proposed a new board for ASI, which is currently under criminal investigation for mismanagement of funds. The board will be headed by Giorgio Fiocco, an atmospheric physicist, and will take over from a temporary management team put in charge last September to sort out the chaos in the agency.

If approved by Parliament, the proposed

board offers a relatively fresh start for ASI; only one of its twelve members has served before. Parliament has until 15 March to raise objections; otherwise the board will automatically be approved. The timing of the announcement, made when Parliament is not sitting because of election campaigning, means that the board is likely to be approved by default.

Whatever the outcome, ASI's troubles continue. On Monday 28 February, police confiscated files from the agency in preparation for a criminal investigation into the alleged mismanagement of billions of lire.

Alison Abbott

Bulgarian strike lifts science budget

Munich. A two week "symbolic strike" by Bulgarian scientists, in which the scientists turned up at their laboratories but did not carry out any work, resulted in partial victory last week when the government agreed to increase this year's budget for the Academy of Sciences by over six per cent.

But the government is still rejecting the scientists' second demand, namely a new science law which would establish a science policy agenda with a guaranteed level of financial support.

Last year around, 700 million leva was allocated to the Bulgarian Academy of Sciences to run its 75 research institutes, comprising over 10,000 staff (of which 4,300 are scientists). Because of the government's financial difficulties, however, only 642 million leva were in fact distributed, barely enough to keep the institutes alive. (All subscriptions to foreign journals, for example, were stopped last year).

The situation was made worse by the

announcement last December that a new technocrat government installed by the Bulgarian president at the end of 1992 planned to give the academy only 930 million leva (US\$23 million) this year. With inflation running at nearly 50 per cent, this would have meant a considerable fall in funding in real terms.

The announcement initiated a campaign last December by the Academy of Sciences for more money and for a national science policy which would establish science as a priority for the country's economic recovery. The campaign culminated in the strike, which started on 14 February. The government ended the strike two weeks later by offering an additional 60 million leva to the academy's budget.

The scientists' strike came on top of a much longer strike by university students, which began in late January, for more money for education. This strike was also called off last week because students decided to sit their winter semester exams.

The government has, however, agreed to improve the welfare and living conditions for students from the countryside. Since the 1970s, foreign students had been given priority treatment on issues such as housing in the nation's capital, Sofia, to the cost of Bulgaria's own students.

Naum Yakimoff, general secretary of the Academy of Sciences, says that neither the scientists' nor the students' strikes were entirely about money. Equally important, he says, is the need for science to be given its proper recognition if the country is to pull itself out of its dire economic situation.

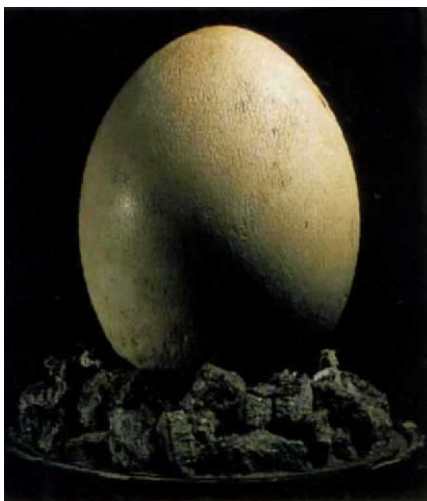
Yakimoff is continuing to press for a law defining a science policy for Bulgaria, which would include a guarantee of a minimal level of financial support pegged to the country's gross national product (GNP). He would like to see Bulgaria spend 2 per cent of GNP on science.

Alison Abbott

Elephantine egg heads for the hammer

This egg of the extinct flightless bird *Aepyornis* from Madagascar is expected to fetch over £30,000 when it comes up for sale at London auctioneer Bonhams on 8 April. At up to 33 centimetres long, *Aepyornis* eggs are the largest eggs ever laid, and are big enough to hold the contents of 7 ostrich eggs or 180 hen eggs.

Aepyornis and its kin — the elephant birds — resembled heavyweight ostriches standing up to about 3 metres high on massive legs and weighing up to 1,000 pounds (450 kg), three times as much as an ostrich. Elephant bird fossils have only been found on Madagascar. The last living *Aepyornis* was reported in 1658. Only about 30 unbroken *Aepyornis* eggs are known to exist, most of them in public collections.



Biosphere 2 experiments with scientific visits

Washington. As a new team of inhabitants re-occupied the huge, hermetically sealed Arizona greenhouse known as Biosphere 2 this week, there are signs that the rift between the US scientific establishment and the private sponsors of the \$150 million facility may be healing.

Six so-called Biospherians entered the three-hectare attempt to replicate the global ecosystem on Sunday, and will spend between four months and a year under glass. But in contrast to the initial two-year occupation of the facility which ended last September, the second stint of experimentation will include the arrival and departure of outside scientists for shorter periods in order to work and collect data.

The change in strategy has boosted hopes among the supporters of Biosphere 2 that some of the rigorous science that its detractors have accused it of neglecting will now be done.

Publication of work from the first phase of inhabitation — including scientific papers on oxygen loss by Wallace Broecker of the Lamont-Doherty Earth Observatory and others (see *EOS* 75, 33; 1994), and on health effects by Roy Walford of the University of California at Los Angeles (see *Proceedings of the National Academy of Sciences* 89, 11533; 1992) — has bolstered the credibility of the facility.

"I guess that gives us some sort of stamp of approval," says Jack Corliss, a former oceanographer who moved from doing computer modelling work on evolution at NASA to become Biosphere 2's research director last March. As Corliss arrived, the facility's outside advisory panel disbanded itself, citing various differences with Space Biospheres Ventures (SBV), the private company which owns and operates the experiment.

Biosphere 2's many critics, including members of the defunct advisory panel, say they hope that the more flexible structure of the new phase of operation will increase the usefulness of the facility. But they remain scathing about the organization of the project.

"I still don't think they are doing any science," says former panel member Gerald Soffen, director of university programmes at NASA's Goddard Space Flight Center in Maryland. "The design and construction are great, but it hasn't proven itself." But Soffen says he hopes that outside scientists will now get involved in exploiting the "terrific capability" of the facility.

Paid for mainly by Texan billionaire Edward Bass, Biosphere 2 attempted to house eight people in almost complete isolation. It was widely ridiculed when one inhabitant had to leave for medical treatment, and when special equipment, such as carbon

dioxide scrubbers, had to be added to sustain the experiment. More seriously, critics charged that before the arrival of Corliss the programme lacked a research director or a serious scientific agenda.

But the project, whose name alludes to the Earth as Biosphere 1, has also become a symbol for those in the environmental sci-

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Biosphere 2 hopes to answer the critics.

ence community who claim there are shortcomings in traditional scientific methods.

"The average scientist is a reductionist," says Eugene Odum of the Institute of Ecology at the University of Georgia, who also served on the advisory panel. Odum says

Warning on cancer gene screening

San Francisco. The National Institutes of Health's (NIH's) National Advisory Council for Human Genome Research is due to issue a warning this week that genetic screening tests for susceptibility to cancer should not become widely used before adequate research has been carried out on their utility, as well as their likely impact on patients.

The warning comes in a statement published in a leading medical journal, urging physicians to avoid using such tests — apart from small controlled research studies — until issues of quality assurance, counselling, medical implications and the dangers of genetic discrimination can be resolved.

The statement is likely to call for coordinated research on such questions, and NIH has already issued a request for proposals for research in such fields.

The council's statement is being made partly in response to an announcement in December by Oncor Inc., a Maryland biotechnology company, that it plans to begin offering linkage analysis for markers associated with a gene implicated in about 10 per cent of breast cancers, as well as in some ovarian cancers.

Oncor has come under criticism from geneticists who argue that no test will be able to provide reliable information about susceptibility to breast cancer until the gene involved has been cloned.

But Doug Dolginow, president of Oncor

that, by looking at a complete ecosystem, Biosphere 2 "has a top-down approach which is the ecological approach. It is as much a human experiment as a scientific one."

The original Biospherians, all of whom were nonscientists, were "resourceful individuals with survival skills," says Odum. "Scientists wouldn't have worked in there; they would be so cantankerous they'd end up killing each other."

Corliss says he is "fairly confident" that the new team has moved along the path of making the project ecologically independent, and that the experiment can now move on to a more flexible regime that will open it up to more visiting scientists.

There will be no new advisory panel. According to Corliss, "We don't need an advisory panel, we have advisory scientists." Such panels are all right for government-funded work, he says, but his project represents "a more entrepreneurial way of organizing things and doing research." He suggests that NASA would have required ten times as much money to run the Biosphere 2 project, "and it still wouldn't be built yet."

Colin Macilwain

subsidiary OncorMed Inc., defends the company's plans by saying that it intends to carry out the analysis for only a small number of families with a confirmed hereditary pattern of cancer.

Each family will have been through counselling to ensure that they understand the significance of the test, he says, adding that the company also has a genetic screening test for colon cancer which it is only offering to patients in research studies until more is known about the effects of the gene.

According to Dolginow, Oncor agrees with the main thrust of the genome council's warning about cancer screening tests, and accepts that a significant amount of further research is necessary before genetic tests become commercially feasible. But it also believes that enough data already exist about the genetic linkage of breast cancer to make its analysis reliable.

David Galas, who was responsible for the US portion of the \$3 billion Human Genome Project until late last year, and is currently vice-president of research at Darwin Molecular Corp. in Washington state, says that such tests need to be introduced carefully, as they challenge assumptions about civil rights, privacy and individual responsibility. In California, a bill that would prohibit genetic discrimination by health insurers has passed the state senate and is awaiting a committee hearing in the state assembly.

Sally Lehrmann

Protests challenge INSERM decision on Benveniste lab

Paris. Twelve prominent French intellectuals last week publicly protested at the decision of France's biomedical research organization (INSERM) to close the laboratory of Jacques Benveniste, the researcher who claims to have shown that extreme dilutions of an antibody solution can retain biological activity.

In a letter to the newspaper *Le Monde*, the protestors, defending the respect traditionally given to *l'idée* among the French intelligentsia, claim that INSERM has discriminated against Benveniste because of his "heretical" ideas. They say that they do not want to enter the "scientific debate", but are keen to defend "freedom of thought" against dogma and the "deification of reason until it becomes folly".

In reply, INSERM says it used routine procedures in reaching its decision not to continue its support of Benveniste's laboratory. Under the organization's rules, a research director must reapply every 12 years to secure the continued existence of his laboratory. But INSERM says Benveniste's laboratory did not even meet the minimum criteria for being re-evaluated, as it failed to meet the requirement that each laboratory have either at least four researchers — including two from INSERM — or else three INSERM researchers.

Benveniste claims he was unable to meet this requirement because INSERM had prevented researchers from joining his laboratory. In particular, he says that INSERM's scientific council recommended against providing a studentship to a young researcher Yolène Thomas last year, on the grounds that his research on the effects of molecules at high dilution "had not been confirmed".

But Claude Amiel, the president of the council, denies that the council tried to prevent Thomas or anyone else joining Benveniste's laboratory.

Benveniste also claims that three of his former students were appointed to research posts by INSERM — but in other laboratories. Only one of these, however, applied to Benveniste's laboratory, and after a first unsuccessful attempt, he was then given a post in another laboratory. The other two students applied directly to work in laboratories other than Benveniste's.

Amiel says that he is sympathetic to the claim that Benveniste's research "could be true". But he adds that, even if the results cannot yet be explained, they should at least be confirmed by others.

Amiel also points out that Benveniste's laboratory will not be closed until June 1995, and that the researchers in the laboratory "still have over one year to prove their theory".

Declan Butler

Top 'research' universities increase their funding lead

London. Further evidence of the gradual emergence of elite 'research universities' in Britain was provided by the announcement last week of the budget allocations by the Higher Education Funding Council for England (HEFCE) for the year 1994-95. (The funding councils for Scotland and Wales allocate money to universities separately.)

In an overall settlement which provides English universities with an average of 4.5 per cent more than they received in the current financial year, the council says it has managed to maintain research funding virtually level in real terms.

A total of £626 million is being provided for research, out of a total budget of £2.61 billion. This is a four per cent increase on last year and, says the council, is in line with the anticipated rise in the retail price index.

But the Committee of Vice Chancellors and Principals (CVCP) is less optimistic; it argues that inflation in research costs — calculated from the university pay and prices index — will be between seven and eight per cent. "Our view is that the inflation forecast is optimistic," says Ted Nield, spokesman for the CVCP.

A statement issued by the CVCP expressed its "disappointment" at the final settlement. "The absence of any real-terms increase does not live up to expectations generated by the recent white paper on science and technology," it said.

The bulk of the £626 million spent on university research by the Department for Education will, as in the past, go to the so-called 'old' universities (in contrast to the 'new' universities, recently upgraded from polytechnics or technical colleges).

Some of the best established are those receiving the greatest increase in research funds as a result of their strong performance in a research assessment exercise carried out

by the HEFCE, the results of which were published last May (see *Nature* 362, 4; 1993).

The current financial year is the first in which research funding through the HEFCE was calculated on the basis of these research assessments. It also led to 19 universities with excellent research assessments being "capped", and the money released in this way being redistributed to those who would otherwise have suffered a major decline in income because of their poor performance. For the next financial year, however, only two universities — Oxford and Cambridge — have been capped.

One result of the HEFCE's funding formula is that some of the biggest increases have gone to universities with the strongest research traditions, in particular Oxford, Cambridge and University College London. "There has always been a top ten or twelve universities who have always received most of the money for research, although it may be a bit more transparent now," says Nield.

But those universities that have been capped complain that, without such a redistribution, they would have done even better. "We are unhappy that the research cap is still in place," says Joanna Womack, the finance director of the University of Cambridge. "It has the effect that we lose £3.1 million which we would have received."

The HEFCE hopes that this will be the last year that capping will be needed. Indeed Derek Roberts, the provost of University College London, which was capped last year but has escaped this year, says that the HEFCE should go further and increase the differential between universities so that it provides even more research money to those at the top end of the assessment scale.

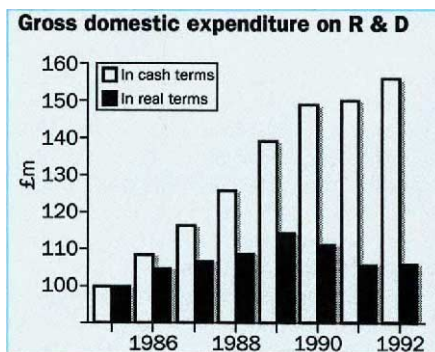
Fiona Gammie

Decline halted in UK on R&D spending

London. The sharp decline in Britain's spending on research and development at the beginning of the 1990s seems to have halted, according to figures released this week by the Central Statistical Office (CSO).

The figures show that, in real terms, expenditure on R&D stayed almost constant between 1991 and 1992, with a decrease in government funding being almost exactly matched by an increase in spending by the private sector. This was in contrast to a five per cent drop in real terms experienced between 1990 and 1991.

The CSO figures also show that, as in previous years, the most rapid increase in R&D funding has come from private non-profit



organizations — mainly research charities — whose support grew by almost 10 per cent in real terms in 1992 to £376 million.

Budget cuts mean no new projects for Indian science

New Delhi. The Indian government appears to have ignored pleas from leading scientists for a large increase in science funding by allocating less to research and development (R&D) next year than needed to keep up with the anticipated rate of inflation. Most of the government's scientific agencies will, for the third year in succession, only have enough money to pay the salaries and housekeeping bills needed to keep existing research projects going, and will be unable to finance any major new projects.

According to figures announced last week, the total allocation for civilian R&D in the 1994-95 budget will be \$1,160 million. This is an increase of only 3.6 per cent (\$43 million) over this year's figure, and compares to expected inflation of 8.5 per cent. Military research will do slightly better; it has been earmarked to receive \$246 million, up 7.5 per cent.

Apart from approving two new research vessels for the Department of Ocean Development, the government has not sanctioned a major new project in science and technology in any of its 400-odd research institutions. At the same time, however, the government is attempting to minimize the impact of 'subcritical' state funding for R&D by setting up two new schemes for mobilizing resources from the private sector.

Under the first of these, companies who import technology from abroad are required to contribute a sum equivalent to five per cent on their royalty payments into the newly launched fund for technology development and application. This will be used to help "indigenously developed technologies reach the stage of commercialization".

Government officials hope that the new fund will generate between \$25 and \$30 million in 1994-95, and that the amount will grow considerably in future years with the significant increase in royalty payments by Indian companies to foreign organizations, which is expected to follow the new

GATT agreements.

Under the second scheme, the government has offered to give a tax credit of 125 per cent on all private contributions to publicly funded laboratories or universities.

The government's new budget keeps R&D support for most industrial sectors at the same level -- or slightly higher -- as the current year. The exception is a marginal reduction in support for the electronics industry; given this cut, the Department of Electronics is to focus its research efforts on promoting projects in photonics and "massively parallel processing machines".

The biggest impact of the funding restrictions is likely to be felt by the Nuclear Power Corporation (NPC), whose plans have already been cut back, and which needs \$380 million during 1994-95 for its new atomic power plants.

The government is planning to provide the NPC with US\$60 million in next year's budget, and has asked the corporation to raise the rest of the money on the open market. But last year NPC only managed to raise \$45 million through the sale of bonds and equity shares, and the prospects for this year are even bleaker.

In contrast, the Department of Space (DOS) will see its budget grow from \$240 million to \$252 million. This confirms the government's resolve to proceed with the development of the geostationary launch vehicle (GSLV) programme on its own, following Russia's refusal to transfer the cryogenic rocket technology.

But if space researchers are satisfied, medical researchers are less so. One concern is that the priority given to the control of AIDS is leading to a resurgence of other communicable diseases. The new budget earmarks US\$27 million for the control of AIDS, more for all other communicable diseases combined; the outlay on malaria control, for example, is only \$2 million.

K. S. Jayaraman

Pope to appoint bio-ethics committee

Munich. The Vatican has announced that the Pope is to set up a new 'Pontifical Academy for Life' to advise him on bio-medical ethics. The move is part of the Catholic church's contribution to the United Nations' Year of the Family.

The president of the academy will be Jérôme Lejeune, a fervent anti-abortionist who is a research director in France's National Centre for Scientific Research (CNRS) and also chairs the department of fundamental genetics at the Necker-Enfants Malades hospital in Paris.

Lejeune will be assisted by one ecclesi-

astical and five lay counsellors. The academy itself will have 70 scientists as members, to be nominated by the Pope. According to the Vatican, academy members will not necessarily have to be Catholics, but they must be prepared to acknowledge that "life begins at the moment of conception".

The members of the academy will also be required to take an oath confirming their support for "the right to life", and their respect for "the catholic principle which condemns contraception, abortion and euthanasia". □

ANC seeks limits on nuclear power in South Africa

Cape Town. The African National Congress (ANC), widely expected to emerge as the largest political party in South Africa's first all-race Parliament following April's elections, has confirmed its commitment to the Nuclear Non-Proliferation Treaty (NPT). It has also stated its belief that South Africa will not need any more nuclear power stations. There is one nuclear plant already in operation.

Speaking at a conference on nuclear policy held in Cape Town last month, Trevor Manuel, the head of the ANC's department of economic planning, said that the party's position on the NPT was consistent with its long-standing support for the Organization of African Unity's declaration to keep Africa free of nuclear weapons.

The highly enriched uranium from South Africa's seven dismantled nuclear weapons will, however, remain in the country, despite reported opposition from the United States. It seems likely that the ANC will continue to allow the uranium to be used by the Atomic Energy Corporation (AEC) to load its SAFARI reactor. This produces commercial radioisotopes, and also irradiates silicon for the manufacture of power-rectifiers.

If elected to power, the ANC has promised an immediate review of the state's subsidy to the AEC, currently due to rise to R509 (US\$100) million in the next financial year. The corporation has already promised to close its uranium enrichment plant by March 1996. But the manager of the plant, Piet Venter, says this could be done "within months" if necessary. The AEC's R&D programme, in particular its molecular laser enrichment process, will also come under review.

The ANC has said that it will consider continuing support for a fuel fabrication plant to convert uranium oxide (yellowcake) to uranium hexafluoride, provided this can be carried out on a viable scale. The future of the corporation's fuel fabrication plant is closely linked to that of the country's nuclear energy industry, which will be considered by a proposed Energy Policy Council set up to formulate an integrated national energy plan for the country.

But a statement issued by ANC says that, given the current overcapacity in electricity generation in the country and the substantial potential for increasing imports of hydro-power from southern Africa, it is clear that South Africa has "no need for further nuclear power stations". It has also promised to review the future of the country's only existing nuclear power station, situated at Koeberg, 28 kilometres north of Cape Town.

Michael Cherry

France's space efforts come down to Earth

Paris. France's Centre National d'Études Spatiales (CNES), the world's third largest space agency and the driving force behind Europe's space efforts, is changing gear. The shift reflects a change in the goals of its space programme from political to economic competitiveness following the end of the cold war, and a massive surge in French spending on military space programmes.

CNES was set up as a civilian space agency in 1961 to reinforce France's self-styled status as an independent arbitrator between the two superpowers. Later, as the main contributor to the European Space Agency (ESA), France used its space efforts to back its claims to both technical and political leadership of Europe.

CNES's priorities have now changed. Maintaining independent access to space through ESA's French-led Ariane rocket programme remains a top priority. But, according to Jean-Daniel Levi, director-general of CNES, the agency has been too preoccupied in recent years with supporting prestige projects such as the Hermès shuttle and the international space station.

This need not mean that ESA should abandon its plans to participate in the space station. If the United States and Russia continue a global project, Europe will be politically obliged to go along, he says. France could also "hypothetically" afford the project, he says, because it will have completed the most expensive phases of development of Ariane V by 1996.

But Levi warns that when ESA ministers sit down to decide how much support to give the space station next year, France's support will be conditional on the agency not making the mistake of letting the costs of the station "crush" its other activities.

This French pragmatism is relatively new. Gérard Longuet, the industry minister, wants CNES to move away from political goals towards improving the competitiveness of French industry. The 1994 budget for CNES reflects this shift. The total budget has increased by 3.4 per cent to FF8.4 billion (US\$1.42 billion). But funding for space science rises 6.4 per cent to FF1.7 billion; and for applications of space technology by 7.9 per cent to FF1.1 billion.

Levi shares Longuet's thinking. This is one reason he wants to stop the recent escalation in CNES's contributions to ESA by freezing investment in new ESA programmes. France is the only European country to have maintained major national and bilateral space programmes alongside commitments to ESA, and wants to protect the budgets of these from further erosion.

Bilateral agreements are easier to manage than ESA programmes, especially in the operational phase, says Levi. Indeed, the success of the joint French-US altimetric

satellite Topex-Poseidon, and the French-led SPOT Earth observation programmes, has been widely acclaimed. "This is how we see our future now," says Levi "But we will stay good Europeans."

In principle, CNES is well placed to increase its emphasis on developing the applications of space-based technology, such as Earth observation techniques and navigation systems. A showcase of the Colbertian



Levi: need for new thinking.

model of creating technology-based industries through massive public investment and planning, CNES — which has big stakes in four major companies — is the only national space agency in Europe able to take projects from conception to operation.

The potential economic impact of Earth observation, France's main national programme, is greater in the long-term than, for example, construction of infrastructure, the impact of which goes little further than the value of the construction contracts. A rise in the sale of images from SPOT, for example, creates a threefold increase in the market for the systems to handle them, and a tenfold increase in the market for the interpretation and use of the data.

But while the Arianespace company has been a success, it is not clear whether CNES can repeat this in creating new industries far downstream from its traditional areas of competence. Japan, for example, lags be-

hind Europe in satellite design, but has become the world leader in ground stations for telecommunications satellites.

Levi also says that the FF485 million CNES will spend this year on 'advanced concepts' research such as artificial intelligence, robotics and new launch systems is not enough to keep pace with the "increasing ferocity" of international competition.

A plan to begin a programme of quick and small scientific missions may improve the situation by increasing the turnover of new technologies. In response to requests from researchers, CNES intends to start a programme of such experiments later this year. The time to launch will be reduced to just three years by using small Russian and Ukrainian launchers.

Another major policy change at CNES is a *rapprochement* with the French military. The Gulf war has convinced the army that battlefield superiority now depends on transferring ground-based systems command, control, communication, and intelligence systems into space.

The budget of the military space programme has rocketed from FF697 million in 1987 to over FF4 billion this year; and the French parliament's defence committee recently complained that even this growth is not enough to meet military needs.

CNES has always enjoyed close links with the military, sharing the same industrial base. These links were institutionalized last year when the Ministry of Defence was given shared responsibility for CNES. CNES is now developing a common strategy with the military space programme.

This use of the SPOT-4 platform for developing the Helios spy satellite, for example, has reduced costs by FF1.3 billion. Even with such reductions, France cannot afford to pay for a complete military space system alone. But as Europe is still far from having a common defence policy, it will probably seek to share costs by pursuing individual bilateral and multilateral programmes.

Declan Butler

Wanted: good home for accelerator components

Washington. The Department of Energy (DOE) and the state of Texas have asked for suggestions on what can be done with the various bits and pieces of the defunct Superconducting Super Collider (SSC).

The remnants of \$2 billion of public investment on the abandoned particle accelerator include "buildings for the linear accelerator as well as its negative hydrogen ion source and 2.5 million electron volt first stage (the higher energy sections are not complete); nearly 15 miles of tunnel; and an integrated network of distributed computer workstations", says a note from the DoE.

Other items for which a new use is being sought include a machine shop, half a million square feet of office space, and well-equipped magnet laboratories. Aspiring SSC carpetbaggers are invited to meet at Waxahachie on 22 March, with expressions of interest due in writing by 15 April.

As yet, this is not an asset sale. The DoE wants proposals for uses that will be good for science and — particularly pertinent to the impending legal battle between the federal government and Texas over the \$400 million which the state put into the project — good for Texas.

Colin Macilwain

How to be a centre of excellence

SIR — John Maddox (*Nature* 366, 717; 1993) presents four basic reasons for the success of the Japanese Institute for Space and Astronautical Science (ISAS). I was reminded immediately of a British research institute which up to about 20 years ago fitted his criterion quite closely. It was (1) autonomous and tightly run, (2) relied on science and engineering equally for its success, (3) was human in scale (100–150 staff) and (4) staff of all grades worked closely together to get the practical work done. This institute was called the National Institute of Oceanography (NIO). The neutrally buoyant deep-ocean float and the GLORIA side-scan survey instrument were among its developments. In the 1980s, under contract to the US Geological Survey, GLORIA was used successfully to survey an area roughly equal to that of the continental United States with a down-time from all causes (mainly bad weather) of less than 5 per cent. After 20 years, this instrument is still capable of regularly producing surprising discoveries, and its development continues.

In 1960, about two layers of administration were needed for the director of the NIO to obtain support to build the research vessel *RSS Discovery*. In Britain today, a decision of similar relative size may require nine or more layers. This increase in complexity has been caused in the main by an entirely laudable desire to ensure accountability for public money spent on research.

Instead, it has diffused accountability, and tends to turn the efforts of good people at all levels away from active science and into administration. The situation cries out for flatter structures, with real authority and accountability delegated as closely as possible to the researcher at the bench or in the field.

An active university research supervisor can train in his working lifetime many times the number of new researchers required to maintain a steady-state population. This may not always be appropriate. When this is the case, or when a field of research requires long-term planning or has relatively high infrastructure costs, the research institute can be an appropriate way to meet these special needs. However, research institutes cannot flourish without close university contacts. The late Roger Revelle, who was a great director of the Scripps Institute of Oceanography, is recorded as having said that a marine institute not linked to a university could expect to survive only one generation. He would be pleased that the Institute of Oceanographic Sciences (successor to the NIO) is now being integrated into a much larger marine centre at the University of Southampton. This development, along with

similar changes occurring at other marine institutes, should be of considerable benefit to British marine science. Its management structure is now being determined. The lessons Maddox draws from the successful operation of just such a university-based research institute in Japan could be applied in Britain to very useful effect. To recapitulate, these include a really meaningful autonomy, a close integration of pure science with applied disciplines (especially engineering) and an open human-scale community with a flat structure allowing active involvement in research at all levels of the organization.

T. Roger S. Wilson

*Firbank,
Portsmouth Road,
Milford,
Godalming, Surrey GU8 5DR, UK*

Of sound mind

SIR — J. R. Smythies¹ states that Canon C. E. Raven told him that Wittgenstein “suffered from paranoid delusions”. Smythies then elaborates on a “speech disorder known as schizophreneze”. Smythies says that he attend[ed] the weekly meetings of Wittgenstein’s disciples and found the “thoughts produced” by this group “very like the thoughts and mode of thinking that troubled my schizophrenic patients”.

On 12 November 1993, we interviewed Professor Norman Moore about Smythies’ claims². Moore was professor of psychiatry at Trinity College, Dublin, and the leading Irish psychiatric clinician of his day. In the late 1940s, Wittgenstein had been referred to Moore for ‘another opinion’ by Dr Maurice Drury, a psychiatrist working at St Patrick’s Hospital in Dublin, and a friend of Wittgenstein. Drury, according to Moore, “was worried” about Wittgenstein, who was on one of his extended visits to Ireland. Unlike Smythies, who had “no personal contact with Wittgenstein”, Moore saw Wittgenstein about five or six times in St Patrick’s Hospital. Moore categorically stated to us that Wittgenstein was not a schizophrenic. Moore described Wittgenstein when he saw him as a “depressed and sad man”, who was “down with depressed affect” and “gloomy”, that he spoke “slowly” and was “slowed down”.

There is extensive evidence of Wittgenstein’s depressive moods, as is clear from Ray Monk’s careful biographical study of Wittgenstein. Thus Monk³ quotes Wittgenstein writing to Rush Rhees: “First suffered terrible depressions.” Later Monk³ talks about Wittgenstein’s “depression as Christmas approaches”. There

is no doubt that Wittgenstein was a sensitive, highly strung personality, whose interpersonal relations could be difficult; and indeed Moore describes him as such. But Moore was certain that Wittgenstein showed no signs of schizophrenia. We think this evidence is important, as Moore seems to have been the only psychiatrist to interview and perhaps treat Wittgenstein in a professional capacity.

Smythies’ evidence that Wittgenstein suffered from thought disorder or paranoid delusions amounts to little more than gossip or hearsay. The only person who was schizophrenic in the inner Wittgenstein circle, according to Monk³, was Yorick Smythies, who “suffered from paranoid schizophrenia and became a patient of Maurice Drury. He died in tragic circumstances in 1981.”

Michael Fitzgerald

David Berman

*Department of Psychiatry,
Department of Philosophy,
Trinity College,
Dublin, Ireland*

1. Smythies, J. R. *Nature* 350, 9 (1991).

2. Fitzgerald, M. & Berman, D. Taped interview with Norman Moore (Dublin, 12 November 1993).

3. Monk, R. *The Duty of Genius* (Jonathan Cape, London, 1990).

Creative budgeting

SIR — In David Dickson’s report “UK councils face £7.5 million efficiency cuts” (*Nature* 367, 501; 1993) he refers to “new money” provided by the Treasury.

Good news? Well perhaps, but we should be careful to distinguish ‘spin’ from fact in government presentations. The spending plans announced in November 1992 foresaw a budget for science staying approximately constant in real terms until 1994–95 and then falling in 1995–96. In November 1993, Save British Science welcomed the budget statement demonstrating that William Waldegrave’s promise to maintain the value of the science budget had been kept, and extended to 1995–96.

There is no “new” money. After the November 1993 statement, a provisional, but unannounced, carve-up between the research councils was made in which £9 million was left unallocated. Of this sum, £7.5 million (leaving £1.5 million for next month’s ‘new money’?) has been added to John Cadogan’s ‘efficiency savings’ of £7.9 million (so far unidentified) to make the £15.4 million in question.

John Major, in his ‘back to basics’ speech on science, has promised a ‘real’ increase for science next year. Had we seen the small real increase promised in 1991, science would already be about £40 million a year better off.

J. H. Mulvey

*Save British Science Society,
Box 241,
Oxford OX1 3QQ, UK*

Razor-blade of life

SIR — Occam's razor is often justified on the grounds that our world is simple, as proved by fundamental physics. Equally often, Occam is declared wrong on the grounds that there is complexity in the world, for instance in biology. In both cases, Occam's razor is interpreted as a statement about the world, rather than a principle of scientific method. In his compelling discussion of Hume's problem (*Nature* 366, 105–106; 1993), Allan Goddard Lindh reopens this issue. But my view is that the interpretation of Occam's razor as a statement about the world is of no interest to the practising scientist. This argument does not involve any new philosophical idea.

A careful reading of Occam's statements "Plurality is not to be assumed without necessity" and "What can be done with fewer assumptions is done in vain with more", shows that Occam's principle is independent of the nature of the world: in any possible universe, we should seek the simplest feasible interpretation of our observations, even though this interpretation may be very complex in absolute terms.

As an example, a double helix is more complex than a single helix, but it is still the simplest structure that can account for what is known about DNA and heredity. No doubt an explanation involving a more complex structure could be contrived, but it is not necessary.

As Karl Popper pointed out in *The Logic of Scientific Discovery*, Occam's razor is a consequence of the requirement that hypotheses be falsifiable: if a theory contains fewer parameters than the facts that it can predict, then the theory has already been corroborated.

For instance, if ten points on a graph lie on a straight line, then the straight-line hypothesis could have been falsified eight times, but was not. In this case, a falsificationist application of Occam's razor requires that we accept the straight line as a working hypothesis, extrapolate from it, and test its predictions; it does not require accepting the straight line uncritically.

There are two other methodological reasons to seek simplicity: simple models are easier to simulate and they give more insight into the processes that one is trying to understand. However, I believe that the greater falsifiability of simpler hypotheses is the main reason why, in any possible world, we can do no better than trying out the simplest hypothesis first, independently of whether it is more likely to be true.

Arthur Pece

Department of Visual Science,
Institute of Ophthalmology,
Bath Street,
London EC1V 9EL, UK

SIR — It is not surprising that Lindh concludes that Popper did not really succeed in solving Hume's problem of induction. Given that falsification is merely another (that is, negative) form of generalization of empirical observations, its methodological justification must undergo the same severe criticism as has already been formulated for its logical counterpart (that is, verification).

But it is surprising that Lindh totally neglects the conceptual starting point of all of Popper's well-known objections to inductive procedures. A closer look at the totality of Popper's work clearly reveals that his main concern was regularly concentrated on the categorical rejection of the commonly accepted idea that intelligent organisms are instructed about the environment by a simple assimilation of external information. This point is much more important than the purely methodological controversy. It leads directly to the fundamental problem of the origin of any kind of knowledge and, by doing so, goes far beyond the subordinate question about the functional relationships between direct perception and (always conjectural) thinking.

Modern physics, however, is probably not the scientific discipline to provide appropriate answers and to solve this basic problem. Rather, it is the task of biology to explain how living systems became able to transform 'senseless' physical information into (within the system) meaningful knowledge. At the time of Hume, it was not yet possible to imagine a phylogenetic evolution of human rationality itself, but today this is an idea strongly suggested by both the empirical and theoretical results of comparative ethology and evolutionary epistemology. Popper, by adding the term 'evolutionary' to his theory of *Objective Knowledge* (1972), did nothing more than to address exactly this issue.

Adolf Heschl

Konrad Lorenz Institut,
A-3422 Altenberg/Donau,
Adolf-Lorenz-Gasse 2, Austria

Don't call us . . .

SIR — Most journals now promise rapid review and publication, but, I wondered, has this recent emphasis on speed made any difference? Accordingly, I determined the time required for review and publication of papers that I submitted to peer review from 1977 to 1993.

Average reviewing time ($\pm$ s.d.) stayed constant at 8.7 ± 2.3 weeks in 1977–84 ($n = 9$), 8.7 ± 2.7 weeks in 1985–89 ($n = 10$), and 9.2 ± 3.0 weeks in 1990–93 ($n = 8$). I defined reviewing time as the interval between the date the editor received the

paper and the date on the letter enclosing the reviews. The fastest review, 4 weeks, was in 1987, and the slowest, 15 weeks, in 1988. Interestingly, much of this time was not used by reviewers. Papers commonly spent several weeks on an editor's desk both before being sent out for review and after the reviews were returned. (Editors' assistants seldom cover for their bosses.) On the other hand, the average time between acceptance and publication decreased from 18.7 weeks in 1977–84 to 14.2 weeks in 1990–93, ranging from 8 weeks, in 1978, to 35 weeks, in 1984.

If my experience is general, then 8 or 9 weeks may represent a practical limit that balances the patience of authors with the workloads of editors and referees. Given this seemingly immutable barrier, I plan, though cannot promise, to leave editors in peace for a couple of months after submitting a paper. Should any journal actually complete a review within the promised 2–3 weeks, I shall, of course, be delighted.

Jeffrey Boone Miller

Neuromuscular Laboratory,
Massachusetts General Hospital,
Charlestown, Massachusetts 02129, USA

Food for worms

SIR — As a formulation chemist, I am interested in a recent notion put forward by Daedalus (*Nature* 366, 20; 1993) regarding sustained release of pharmaceuticals by intestinal worms. By coupling the production of the pharmaceutical to some internal clock of the worms, pulsatile drug release could be obtained, which for some treatments is more advantageous than steady release. But the increased cost of feeding all these worms human cuisine would far outweigh the benefits obtained by steady dosing.

Actually, the best use of such worms would be in the control of human eating disorders, a fact that Daedalus all too quickly glosses over. Specifically developed tapeworms could be constructed for the over-indulgent and such disorders as anorexia nervosa and bulimia would disappear. Cleavage sites engineered into different sections of the worm could be actuated by various orally active enzymes that would tailor the worm's length and, therefore, its metabolism to provide an appropriate level of stasis for any individual. Any increased food consumption (no doubt some people will exploit the system with ever increasing times at the feed-bag) would be more than compensated with the millions saved by dispensing with fitness centres, exercise machines and expensive fly-by-night diet plans. It would appear that Daedalus missed the boat on this one.

David F. Johnson

557A South Dove Road,
Yardley, Pennsylvania 19067, USA

Next step for climate-change analysis

Bert Bolin

In this article, the chairman of the Intergovernmental Panel on Climate Change, explains the preparations for the organization's second assessment, to be published in September next year.

THE Intergovernmental Panel on Climate Change (IPCC) is responsible for the scientific and technical assessment underlying the Framework Convention for Climate Change agreed in June 1992 in Rio de Janeiro. IPCC was created in 1988 by the United Nations Environmental Programme (UNEP) and the World Meteorological Organization (WMO), and its work is partly financed by these organizations and partly by voluntary contributions from various countries.

IPCC's first assessment* in 1990, updated in 1992, analysed the scientific basis for the likelihood and character of climate change during the next century from increasing emissions of greenhouse gases; the likely impact of any change in climate; and the strategies that should be devised to adapt to any climate change. The future form of cooperation between IPCC and the convention will be decided at the first meeting of the parties to the convention, to be held next March in Berlin.

Our task is to assess knowledge rather than to recommend measures to be taken. We plan to have completed the next phase of our assessment of the available scientific and technical literature by September next year, when we will produce our "Second Assessment". The main task at Berlin will be to establish a basis for assessing how different countries contribute to the increasing concentrations of greenhouse gases in the atmosphere. Countries will have to adopt common methodologies for computing national emissions, for example, and to agree on a standard form for reporting them. The relative importance of different greenhouse gases will need to be agreed.

It does not seem likely that by 1995 negotiations will have progressed very far on how to achieve the prime goal agreed at Rio, the "stabilization of greenhouse gas concentrations in the atmosphere that would prevent dangerous anthropogenic interference with the climate system". Just what is meant by the word "dange-

rous" needs to be sorted out. IPCC will bring together basic knowledge of relevance in this context, but reaching agreement is a political issue and must therefore be achieved through intergovernmental negotiations within the framework of the convention.

To produce our report by September 1995, we will analyse thousands of scientific papers. Specialists wishing to contribute should submit such material to the most relevant of the working groups whose addresses appear at the end of this article. The lead authors for each section of the report will reflect a balance among different points of view, and will include at least one specialist from a developing country. The lead authors of each working group will identify disparities of view for which there is significant scientific or technical support, and will circulate the first (bulk) draft for peer review. The next stage will be review by government specialists, and the final revisions will be used for preparation of three summaries for policy-makers. The Second Assessment will contain about 50 sections, involving more than 200 contributors.

IPCC's conclusions have been criticized in *Nature*, other parts of the scientific literature and, to a greater extent, in the popular press, mainly for lack of openness about uncertainties and for brushing aside controversies. Although we did repeatedly emphasize the uncertainties in our present understanding in our earlier reports, we will make further efforts in our next assessment to present the full picture as carefully as possible. We shall be as open as we can be by widely circulating early drafts of our report.

Certainly, the issue of climate change is complex. But although I respect the views of a few scientists who believe that little can be said about future climate based on our present knowledge, I believe that balanced and careful reviews of current knowledge are of great value both to help public awareness and to assist political decision-making. Policy-makers are more likely to seek scientists' advice if scientists are explicitly addressing controversies in the scientific literature and attempting to clarify their implications.

An increasing number of semi-popular articles about climate change are being published, some predicting imminent catastrophe and others accusing IPCC of exaggerating future dangers. Many of these articles emanate from special-

interest groups, and all are supposedly based on scientific findings. This state of affairs is hardly surprising, and scientists will simply have to pursue good leads in order to arrive gradually at generally agreed interpretations.

More problematic are the popular articles in the daily press, written by people, often well-known in other areas, who do not know much about the basic scientific literature in the field. In my view, it is not possible to resolve key scientific issues in articles aimed at the general public. Although concepts such as global warming or environmental hazards can be easily understood by everybody, the scientific issues cannot so easily be explained. For example, theoretical models of the climate system have been criticized, but any substitute also represents an interpretation of the available information, perhaps using a conceptual model, which is likely to be more primitive than a computer model. Dozens of research groups have devoted hundreds of man-years to models of climate change. It seems to me reasonable that reservations about their results should be supported by thorough analysis. Sadly, too often this is not the case.

As chairman of IPCC, I welcome thorough, critical analyses. But it is essential that such arguments are themselves published so that they can be appraised and used for the IPCC assessment process, thus adding to their credibility. □

Bert Bolin is at Kvarnasvagen 6, 18451 Osterskar, Sweden.

Deadline for contributions to the first draft of the Second Assessment is the end of June 1994. Contributions should be supported by reference to peer-reviewed articles and reports available internationally, and with preprints of unpublished material. Finalization of the first draft, peer review and revisions will proceed during the rest of 1994.

Working group I support unit. Scientific assessment of likely future climate changes. Dr Bruce Callender, Meteorological Office, Hadley Climate Centre, London Road, Bracknell RG12 2ST, UK. Fax +44 344 856912.

Working group II support unit. Impacts of climate change; means for mitigation and adaptation. Dr Richard Moss, 300 D Street, SW, Suite 840, Washington, DC 20024, USA. Fax +1 202 554 7830.

Working group III support unit. Interdisciplinary issues, particularly integrated socioeconomic analyses. Dr Eric Haites, Environment Canada, La Salle Academy, 373 Sussex Drive, Block E, Ottawa K1A 0H3, Canada. Fax +1 416 369 0922.

*IPCC's first assessment is published as:
Houghton, J. T., Jenkins, G. G. & Ephraums, J. J. (eds) *Climate Change Working Group I. Scientific Assessment* (Cambridge Univ. Press, 1990).
Tegart McG. W. J., Sheldon, G. W. & Griffiths, D. C. (eds) *Climate Change Working Group II Impacts Assessment* (Australian Government Publishing Service, Canberra, 1990).
Bernthal, F. M. (ed.) *Climate Change Working Group III. Response Strategies* (US Department of State, Washington, DC, 1990).
Houghton, J. T., Callander B. A. & Varney, S. K. (eds) *Climate Change 1992. Working Group I. Supplementary Report to the IPCC Scientific Assessment* (Cambridge Univ. Press, 1992).

Towards more measurement in biology

Even the best known systems in biology may not be as well-understood as is generally believed, which means that understanding is incomplete and may even be misplaced.

WHAT does one do when challenged to justify in public* what *Nature*, or at least this small part of it, has been preaching about the need for more quantitative observation and analysis in molecular biology? What better preparation could there be than to reread Mark Ptashne's "*A Genetic Switch*" (Cell Press and Blackwell; 1986)? For there is no more vivid account of a deliciously neat device in molecular biology than this tale of how bacteriophage λ infecting an *E. coli* cell can either kill its host by multiplying at its expense or can be incorporated in the host genome, cryptic until an insult is delivered to the cell.

Nothing in the plea for a greater regard for quantitative data is meant to diminish molecular biology. On the contrary, as Ptashne's book shows clearly, the structure of DNA provided not just the answer to the question "What is the physical embodiment of inheritance", but solved the problem of the physical mechanism linking genotype and phenotype. Schrödinger's "*What is Life*", rich though it was in speculation about the nature of genes, said little about the need that genes should determine the physical characteristics of the cells in which they are found. And even Watson and Crick, in their celebrated throw-away line, "It has not escaped our attention...", failed to remark that they had revolutionized biochemistry.

Yet molecular biology remains a largely descriptive science. In the various human genome projects, for example, the objective is essentially to list the genes and to specify their nucleotide sequences, as well as the sequences of the regions of DNA that hold the genes together in the chromosome. The naming of parts does not by itself yield understanding.

None of that implies that the human genome projects will tell us nothing about the working of the cell. On the contrary, there is every likelihood that clever people poring over the listings as they accumulate will learn, simply by inspection, a great deal about the life of the cell. The mere arrangement of regulatory genes must say something important about gene regulation. And it will be invaluable to learn, by the comparison of different genomes, how genetic rearrangement contributes to speciation. And there will be practical benefits, some gene therapy, for example.

So the complaint that molecular biology is insufficiently quantitative at present is a modest and subtle complaint: molecular biology would be in even better condition than it were otherwise. The lysogeny of *E.*

coli is a good illustration because it is such an arresting tale that has been so well described by Ptashne.

The essence of the tale is easily told. Of λ 's 17 genes, one produces a protein called repressor whose effect, when appropriately bound to viral DNA, is to ensure that only repressor remains active. That is the case whether the λ gene is incorporated in the host genome as a linear double-stranded stretch of DNA as in lysogeny or, alternatively, is formed into a circle as when injected into a cell from an external virus particle. Either way, no other λ genes than repressor can function.

The elegance of the mechanism rests on a second λ gene called *Cro* which, if there is no repressor protein present to repress it, will bind to the same DNA sites and repress the activity of repressor. This ying-yang way of doing business is generally taken as a paradigm for the working of switching mechanisms in other gene systems. How well is it explained?

Qualitatively, there are no problems. In lysogenized cells, there is repressor protein to switch off all but repressor in infecting λ DNA, so that viral replication and destruction of the host cell by lysis do not follow. The difficulties are quantitative. But first one needs to know that the functional form of the repressor protein is a dimer of the native molecule, and that repressor dimers can bind to three contiguous sites on λ , arranged in order of decreasing affinity. When bound to the most attractive site, the protein increases the affinity of the second site and their twin occupancy prevents transcription of all viral genes except repressor itself. But if the third site, with the lowest affinity, is occupied, repressor production ceases. It is reckoned that the saturation of all three sites occurs when there are 100 or so free molecules in the cell.

What would a physical chemist make of this? The first observation would be that there is a cascade of chemical equilibria: that between the monomers and the dimers, then between cytoplasmic dimers and the three successive stages of occupancy of the DNA binding sites. On the face of things, the law of mass action applies. The concentration of dimers will be proportional to the square of the concentration of monomers, and similar arguments will apply to the occupancy of the three genomic sites. But is that likely to apply when there are only 100 or so molecules?

That is where the difficulties arise. Even to assess whether mass action applies re-

quires consideration of such things as the diffusion rates of repressor molecules in the cytosol and probably, in the last resort, a model for the distribution of ribosomes (where protein molecules are synthesized) relative to the three active sites. But so far as can be told, these assessments have not been made, nor has there been a serious attempt to calculate the likely fluctuations of the performance of 100 molecules of repressor protein — which may be vital for an understanding of the function of the genetic switch (not to mention the likelihood that the cell concerned will escape lysis).

Indeed, the chances are that this delicate genetic switch is not a phenomenon of equilibrium thermodynamics at all, but a kinetic one. And if that is possible in this rudimentary case, how much more likely is it that, in the functioning of the transcription factors that transcribe genes in eukaryotic cells, kinetic considerations will be dominant? There, the active component is a polymerase molecule, but there are typically at least six others involved, each of which is added to the complex in an apparently precise order. How can that be arranged? Or, to turn the argument around, if the law of mass action applied, the required concentration of the last of these protein components would probably be orders of magnitude greater than that of the first, which is not the case.

This apparently arcane argument is not irrelevant. There is at least one well-known case, that of the chemical composition of interstellar gas, where equilibrium arguments give answers radically different from the measurement. The point is important not merely because we need to know what is really happening, but because, if the behaviour of repressor is determined kinetically and not by equilibrium thermodynamics, all the energetic implications of what appear to be equilibrium constants may be spurious.

But this is not the only source of uneasiness about the state of what must be one of the most elegant of all the tales molecular biology has to tell. There remains the problem of knowing how the molecules of repressor fold into their characteristic dumb-bell shape, why dimers are so much more stable than monomers and how an α helix in the amino-component interacts with DNA at the binding sites. To be more precise, what happens has been determined by elegant genetic experiment; why that, not something else, happens remains to be discovered.

John Maddox

*Gist of a talk at the American Institute of Medical and Biological Engineering, Washington DC, 7 March 1994.

Towards synthetic enzymes

Leonard F. Lindoy

THE design and synthesis of molecules that might mimic the recognition and catalytic behaviour of enzymes is attracting more and more attention from chemists. Last month, Mary McPartlin and co-workers¹ reported the X-ray structure of one such supramolecular system — the tris(pyridine) adduct of the zinc porphyrin cyclic trimer (**1**). Earlier work had shown² that **1** is able to bind two suitable substrates in such a way that reaction between them is enhanced (in parallel to the behaviour of many enzymes). The promise of such research is that it will deliver new chemical processes that are both more energy-efficient and more specific in terms of the products formed than existing processes.

The X-ray structure confirms that there is room for two or three medium-sized guest molecules in the central cavity of **1**, in accord with earlier predictions based on space-filling models and molecular

mechanics calculations. An unusual interpenetration of pairs of cyclic trimers occurs: one porphyrin of each trimer is a guest in the cavity of the other. The result is an interlocked 'dimer of trimers'. The flexibility of the individual cyclic trimers is demonstrated by their marked deviation from 3-fold symmetry in this solid-state structure. For achieving optimum host-guest formation, such flexibility may prove important in allowing the cavity to respond to the steric demands of different substrates.

The authors believe that the observed dimer formation in the solid will not be of much significance in solution; it will have little influence on the ability of **1** to incorporate suitable substrates and hence promote their reaction. Indeed, a preliminary X-ray study shows that similar dimer formation in the solid is absent in the corresponding tris(quinuclidine) adduct, even though it was crystallized under identical conditions to the tris(pyridine) adduct.

In progress towards a 'synthetic enzyme', Sanders and co-workers² have demonstrated that **1** binds the diene **2** and the dienophile **3** in its cavity in such a way as to promote a Diels-Alder reaction between them. The reaction between the bound substrates proceeds at an initial rate some 6,000 times faster than for the equivalent pair in the absence of **1**. As in enzyme reactions, only one stereoisomer (the *exo* form) of the possible *endo* (**4**) and *exo* (**5**) products was detected at any stage during the reaction. In the absence of **1**, a mixture of both forms is produced, although the *exo* form is thermodynamically favoured (and is the sole product at the completion of the reaction).

There is evidence that each of the substrates **2** and **3** bind by means of their pyridyl nitrogens to adjacent zinc atoms of the same molecule of the cyclic trimer **1**. In this manner the respective reactive sites of **2** and **3** are brought close together in the centre of the supramolecular cavity. In the presence of tripyridyltriazine (**6**), which binds extremely strongly to **1** via simultaneous coordination of its pyridyl groups to all three zinc atoms³, no acceleration of the Diels-Alder reaction was observed. In this capacity, **6** seems to take a role similar to that of a classical competitive enzyme inhibitor.

Although **1** accelerates the above Diels-Alder reaction by several orders of magnitude and is stereoselective for the *exo* form, the reaction is stoichiometric rather than catalytic. That is, the Diels-Alder product binds too strongly to **1** to allow effective turnover.

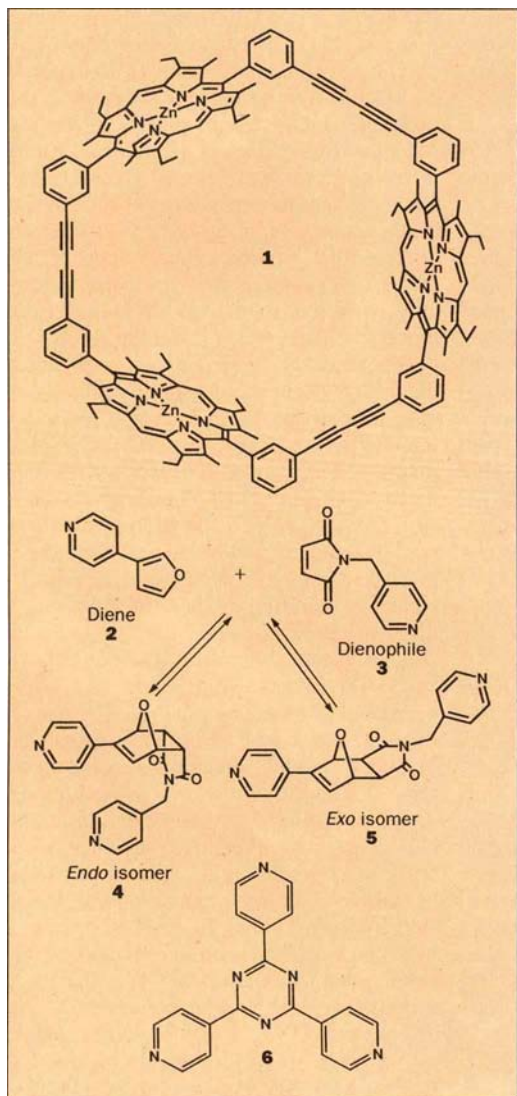
In exciting work soon to be reported by Sanders's group⁴, a related system in which there is an acyl transfer reaction between two substrates bound in the cavity of **1** has been shown to be catalytic. In this case, the catalytic behaviour reflects the fact that only the reaction intermediate is strongly bound in the cavity.

The analogous cyclic dimer and tetramer corresponding to **1** have also been synthesized⁵, giving access to a range of cavity sizes and relative orientations of the substrate binding sites for use in future studies. Clearly, cyclic porphyrin species such as these have the potential to catalyse different reaction types involving substrates of different sizes. The function of the cyclic cavity is mainly concerned with substrate recognition and orientation and the versatility of the overall system is increased through both of these being effectively decoupled from the reaction site. A further refinement of these systems seems feasible to give the prospect of 'tuning' the binding strength of particular guests by incorporating electron-pushing (or electron-pulling) substituents on the outside of the cavity. In this manner more deliberate control of potential catalytic behaviour may be possible.

The synthesis of the cyclic species is interesting in its own right, as higher yields were obtained when the respective ring-closing reactions were done in the presence of suitable template molecules⁴. For the production of **1**, tripyridyltriazine (**6**) was used as the template, because its size and shape are nicely complementary to the cavity in the cyclic trimer.

Although the field of synthetic enzymes remains fairly rudimentary — especially when the extraordinarily high catalytic efficiencies of the natural systems are considered — steady progress is being made. If supramolecular chemistry continues to develop as rapidly as at present, considerably improved synthetic enzymes should emerge in the foreseeable future. □

Leonard F. Lindoy is in the Department of Molecular Sciences, James Cook University, Townsville 4811, Queensland, Australia.



1. Anderson, H. L., Bashall, A., Henrick, K., McPartlin, M. & Sanders, J. K. M. *Angew. Chem. int. Edn engl.* **33**, 429–431 (1994).
2. Walter, C. J., Anderson, H. L. & Sanders, J. K. M. *J. chem. Soc. chem. Commun.* 458–460 (1993).
3. Anderson, H. L. & Sanders, J. K. M. *J. chem. Soc. chem. Commun.* 1714–1715 (1989).
4. Mackay, L. B., Wylie, R. S. & Sanders, J. K. M. *J. Am. chem. Soc.* (in the press).
5. Anderson, S., Anderson, H. L. & Sanders, J. K. M. *Accs chem. Res.* **26**, 469–475 (1993).

Radon risk reassessed

Julian Peto and Sarah Darby

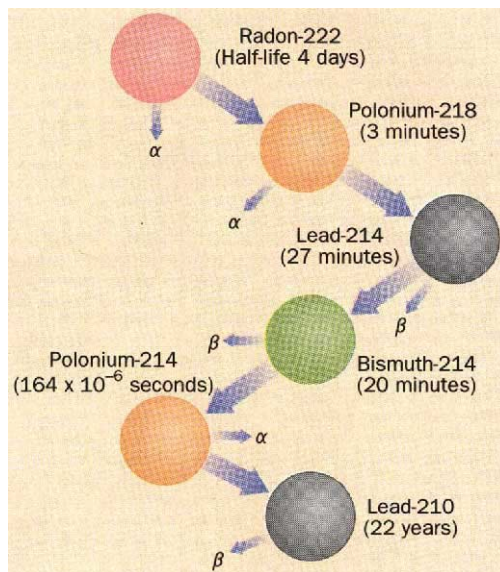
UNDERGROUND miners who are heavily exposed to the radioactive decay products of radon gas (see figure) suffer especially high rates of lung cancer. Extrapolation from these data suggests that radon in people's homes may cause some 2,000 lung cancers each year in Britain¹ and 15,000 in the United States², figures which imply that radon is among the most serious environmental (as distinct from self-inflicted) causes of cancer mortality yet identified in the developed world. It is because of such projections that the International Commission on Radiological Protection now recommends³ an action level for radon in houses in the range 200–600 becquerel (Bq) m⁻³. But there is still a dearth of direct evidence on the effect of domestic radon exposure. An epidemiological study, carried out in Sweden by Pershagen and colleagues⁴, now provides just that. In particular, the authors have gone some way to teasing out the confounding effect of smoking as a cause of lung cancer.

There are considerable uncertainties in the estimates of the lung cancer risk from radon in houses. Even when identical methods of analysis are applied, estimates of risk per unit exposure from the different studies of miners vary by an order of magnitude², and extrapolation from the radon concentrations in mines to the much lower levels in most buildings, and from adult occupational exposure to lifetime domestic exposure, requires a number of assumptions to be made about the effects of age, duration and dose. Then there is the joint effect of radon and smoking. Some studies suggest that the absolute risk of lung cancer from radon in non-smokers is much lower than in smokers⁵; others that the relationship is more nearly additive^{2,6}, with an absolute risk in non-smokers of the same order as that in smokers. This concatenation of uncertainties implies a wide range of plausible risk estimates, particularly for the increasing proportion of the population who do not smoke.

Analyses of the correlation between lung-cancer rates and average radon levels in different regions are unlikely to provide decisive evidence. Several such studies have found a significant positive correlation, but the two largest (of 411 counties in the United States⁷ and of all the counties in England and Wales⁸) found significant negative correlations. Epidemiologists have therefore concluded that the effects of domestic exposure to radon must be estimated from carefully conducted studies of lung-cancer patients and unaffected controls in which detailed smoking and residential histories are

obtained for each individual, together with high-quality radon measurements in both current and previous homes.

Several research projects satisfying these criteria have been published (listed in ref. 9). The results of all but one were generally consistent with the risks predicted from the studies of miners,



The radioactive gas radon-222 is one product of the chain of decay of uranium-238, hence the especial hazard it poses to miners in general and uranium miners in particular. The ensuing decay chain as far as lead-210 is shown, with the alpha and beta particle emissions at each stage and the approximate half-lives of the intermediates.

although the confidence limits were wide, particularly for analyses restricted to non-smokers. The remaining study, which involved 308 women with lung cancer of whom 123 had never smoked, and a similar number of unaffected controls, found no evidence of any radon-related risk¹⁰. This study, although excellent methodologically, may not provide a useful estimate of radon risks in Western populations, because it was carried out in Shenyang in northeastern China where female lung cancer rates are among the highest in the world (due principally to exceptionally high levels of indoor air pollution from coal-burning heating and cooking systems).

The new report by Pershagen *et al.*⁴ is a major addition to the published evidence, particularly for non-smokers. This meticulous case-control study involved 1,360 lung-cancer patients, 178 of whom had never smoked, and 2,847 controls, with radon measurement in all accessible homes occupied since 1947. The resulting overall risk estimate, an increase in rela-

tive risk of 10% for 30 years' occupancy of a home with a radon concentration of 100 Bq m⁻³, is consistent with the central estimate based on the studies of miners, although the 95% confidence interval (1–22%) is wide. For the 43 lifelong non-smokers who were exposed at an average level exceeding 140 Bq m⁻³, the relative risk was 1.5 with an upper confidence limit of about 2.2. Roughly 1 in 250 non-smokers will die of lung cancer, so this upper limit corresponds to a lifetime excess risk due to radon of about 1 in 200.

Non-smokers can thus be assured that their additional risk from prolonged exposure at 200 Bq m⁻³, which is the recommended action level in several countries, including Britain, is likely to be substantially less than 1 in 100, and hence under 1 in 1,000 at 20 Bq m⁻³ (the average level in Britain).

No such reassurance can be offered to smokers, however, whose estimated lifetime risk from radon at 200 Bq m⁻³ from this study has an upper confidence limit of the order of 1 in 10, although the lower confidence limit is close to zero. A Canadian study of 738 lung-cancer cases and 738 controls, now in press¹¹, is unlikely to define the risk to smokers much more precisely, as the relationship between risk and radon level in the various analyses that are presented ranges from weakly positive to weakly negative.

Other work is in progress, and researchers are discussing plans for formal pooling of the various studies. Such meta-analyses require a considerable collaborative effort, as the raw data from each study must be reanalysed by uniform statistical methods, but within a few years these pooled results should provide a much more precise estimate of the risk to both smokers and non-smokers at domestic levels exceeding 200 Bq m⁻³. Such high levels are rare in most countries, however, and most of the collective dose to the population, and hence most of the predicted lung cancer excess, occurs in houses with much lower radon levels, at which predicted relative risks are of the order of 1.1. Conventional epidemiological studies cannot reliably distinguish real effects from bias at such low levels, and it is not clear whether estimates of the total number of lung cancers caused by radon can ever be tested by direct observation.

The resolution of this methodological impasse may come from molecular biology, as heralded by a paper from Taylor and colleagues¹² published earlier this year. These authors looked at tumours in heavily exposed miners and found that about a third of radon-induced lung

cancers carry a specific mutation in the *p53* tumour-suppressor gene, one which is normally very rare (AGG to ATG at codon 249). Such a characteristic marker could refine the precision of excess risk measurement in epidemiological studies by an order of magnitude and provide a direct estimate of the overall proportion of lung cancers caused by radon. Mutations in *p53* characteristic of particular carcinogens have been reported in liver cancers induced by aflatoxin and skin cancers induced by ultraviolet light, and the mutational spectrum of breast cancer suggests that a proportion of cases can be identified which are due to endogenous or environmental mutagens¹³.

It may, however, be premature to assume that lung cancers caused by radon can now be identified by a simple DNA test. Vähäkangas and colleagues¹⁴ did not find this mutation in 19 lung cancers from workers at a different uranium mine, and in a forthcoming letter to *The Lancet*, Venitt and Biggs¹⁵ argue that such a specific mutation is unlikely to be caused by radon. They suggest that a more plausible cause is a mould which is present in certain mines and produces a fungal mutagen structurally similar to aflatoxin. This would imply that a substantial proportion of lung cancers in some studies of miners must be attributed to mycotoxins rather than (or in combination with) radon. The extent to which molecular epidemiology will eventually revolutionize the study of environmental carcinogenesis, and whether the social consequences will be wholly beneficial, remains to be seen. The ability to identify individuals whose cancer was caused by a specific carcinogen is eagerly awaited by lawyers as well as epidemiologists. □

Julian Peto is in the Section of Epidemiology, Institute of Cancer Research, Belmont, Surrey SM2 5NG, UK. Sarah Darby is in the Imperial Cancer Research Fund Unit, University of Oxford, Radcliffe Infirmary, Oxford OX2 6HE, UK.

Clamping down on myosin

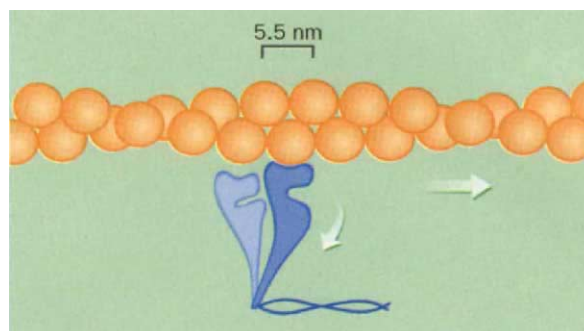
Jonathon Howard

ON page 113 of this issue¹, Finer, Simons and Spudich report the refinement of the *in vitro* motility assay to its molecular limit: they have recorded the movements and forces generated by single (or small numbers of) molecules of the muscle protein myosin. This and related² work on myosin follow fast on the heels of Svoboda

mechanical analogue of the voltage clamp used to study ion channels⁸.

At very low load (in the absence of feedback), Finer *et al.* record transient displacements of 7 to 17 nm with a mean of 11 nm. In the presence of feedback, they record force transients of 1 to 7 pN with a mean of 3.4 pN. Ishijima *et al.*² record transients of similar sizes using flexible glass fibres to apply forces. Perhaps the most exciting finding of Finer *et al.* is their confirmation of the Fenn effect at the molecular level: at high ATP concentration the duration of the force transient is longer than that of the displacement transient, which is precisely what is expected if high load slows down the hydrolysis cycle.

Are the transients due to single molecules of myosin? First, dividing the whole-muscle force by the number



A two-headed myosin molecule interacting with an actin filament. Opening of the ATP-binding pocket would move the filament through 5 nm or so if the base of the head is kept fixed. (After ref. 11.)

and colleagues' recordings^{3,4} from single kinesin molecules. Together, these papers herald a second generation of motility assays with which to probe the mechanism of force production by biological motors with molecular resolution.

Of crucial importance to the study of motor proteins has been the *in vitro* motility assay, a cell-free system in which the movement of purified motors along their associated filaments is reconstituted in the presence of ATP. Such assays formally proved that myosin is sufficient to move actin filaments, and permitted the discovery of new motors such as the organelle motor kinesin. But until recently the motility assay had been limited because the motors were virtually unloaded. To understand how any motor works it is essential to test its performance under different loads; indeed, classic studies by Fenn in 1924 showed that the rate of ATP hydrolysis by a muscle decreases as the load is increased — that is, the engine turns over more slowly at high load⁵.

Several groups have imposed loads on the motors *in vitro*^{6,7}, but none has done it with the elegance and precision of Finer and colleagues' displacement clamp apparatus. The position of the filament is measured with nanometre precision using a photodiode detector; if the motor tries to move the filament, the displacement signal is fed back to a pair of optical tweezers which impose a counteracting force that stops the motion. This counteracting force corresponds to the force exerted by the motor. The apparatus is the

of myosin molecules in parallel gives a force of about 3 pN (ref. 5). Second, rapid length changes of whole muscle reveal a pair (or a small number) of states which are in rapid equilibrium and which are separated by some 10 nm (ref. 9). The transition between these states is the mechanical analogue of the charge movement associated with the gating of ion channels¹⁰ and is thought by some to correspond to the power stroke. Thus the magnitudes of force and displacement transients are compatible with their being due to single molecules.

But there are considerable uncertainties. Most importantly, both the displacements and the forces show a variability that is larger than the measurement noise. This implies that the recordings (see, for example, Figs 2a and 5c on pages 115 and 117) are not from a single, time-invariant molecule nor are they from a number of identical molecules. There are at least two possible explanations. First, the transients may correspond to single molecules but for some reason the molecules are not identical — for example different molecules may have different orientations on the surface. Alternatively, each transient may correspond to a small aggregate containing a variable number of molecules. Quantitative dilution studies or more precise orientation of the motors on the surface are necessary to decide between these possibilities.

The large size of the displacement transients presents a conceptual difficulty. The atomic structure of the myosin head¹¹ suggests that the opening and closing of

1. National Radiological Protection Board Doc. NRPB 1 No. 1 (1990).

2. Lubin, J. H. *et al.* *NIH publ.* 94-3644 (1994).

3. International Commission on Radiological Protection Ann. ICRP 23 No. 2 (1993).

4. Pershagen, G. *et al.* *New Engl. J. Med.* **330**, 159–164 (1994).

5. Commission on the Biological Effects of Ionising Radiation (BEIR IV) *Health Risks of Radon and Other Internally Deposited Alpha-emitters* (NAS/NRC, Washington DC, 1988).

6. Doll, R. *Radiat. Prot. Dosim.* **42**, 149–153 (1992).

7. Cohen, B. L. *Environ. Res.* **53**, 193–220 (1990).

8. Haynes, R. M. *Radiat. Prot. Dosim.* **25**, 93–96 (1988).

9. National Radiological Protection Board Doc. NRPB 4 No. 4 (1993).

10. Blot, W. J. *et al.* *J. natn. Cancer Inst.* **82**, 1025–1030 (1990).

11. Letourneau, E. G. *et al.* *Am. J. Epidemiol.* (in the press).

12. Taylor, J. H. *et al.* *Lancet* **343**, 86–87 (1994).

13. Biggs, P. J. *et al.* *Mutagenesis* **8**, 275–283 (1993).

14. Vähäkangas, K. H. *et al.* *Lancet* **339**, 576–580 (1992).

15. Venitt, S. & Biggs, P. J. *Lancet* (in the press).

the ATP-binding pocket could cause a swinging motion of a long α -helical domain through as much as 5 nm (see figure). But the mean displacement measured by *Finer et al.* is 11 nm, and some motions are as large as 17 nm; such distances seem too large to correspond to a single head, as assumed. Because they are using two-headed molecules, it is possible that *Finer et al.* are actually seeing the strokes of the two heads in rapid succession. Even though the two heads can sequentially bind to one actin filament¹², the binding cannot be identical because the two heads are related by a 2-fold axis of symmetry rather than by translational symmetry. The role of the two heads must be clarified before the whole muscle or single-molecule measurements can be interpreted with an accuracy better than a factor of two.

The displacement clamp apparatus makes many new experiments possible. The possible cooperativity of the two heads can be investigated by comparing the forces of single-headed and double-headed motors. By rapidly clamping the motor to a different position it should be

possible to determine whether the 10-nm force transition characteristic of whole muscle corresponds to the powerstroke. With so many techniques converging on motor proteins — molecular mechanics, enzyme kinetics, molecular engineering and crystallography — we can expect to see great progress in the next year. □

Jonathon Howard is in the Department of Physiology and Biophysics, SJ-40, University of Washington, Seattle, Washington 98195, USA.

1. Finer, J. T., Simmons, R. M. & Spudis, J. A. *Nature* **368**, 113–119 (1994).
2. Ishijima, A. *et al. Biochem. biophys. Res. Commun.* (in the press).
3. Svoboda, K. *et al. Nature* **365**, 721–727 (1993).
4. Howard, J. *Nature* **365**, 696–697 (1993).
5. Bagshaw, C. R. *Muscle Contraction*, 2nd edn (Chapman & Hall, London, 1993).
6. Ishijima, A. *et al. Nature* **352**, 301–306 (1991).
7. Howard, J. *Nature* **364**, 390–391 (1993).
8. Hille, B. *Ionic Channels in Excitable Membranes*, 2nd edn (Sinauer, Sunderland, Massachusetts, 1992).
9. Huxley, A. F. & Simmons, R. M. *Nature* **233**, 533–538 (1971).
10. Armstrong, C. M. & Bezanilla, F. *J. gen. Physiol.* **63**, 533–552 (1974).
11. Rayment, I. *et al. Science* **261**, 58–65 (1993).
12. Hackney, D. D. & Clark, P. K. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5345–5349 (1984).

GAMMA-RAY BURSTERS

Cosmic flasher exposed

Charles Bailyn

GAMMA-ray burst sources (hereafter GRBs) probably constitute the least well understood class of celestial object. Estimates of their intrinsic luminosity vary by over 20 orders of magnitude. The primary impediment to greater understanding has been astronomers' inability to identify radiation other than γ -rays associated with GRBs. Three papers in this issue now reveal a compelling identification of a GRB with an object emitting other kinds of radiation^{1–3}. This, though, is one of a small and unusual subclass, and the observing techniques and models described in these papers will not apply to most GRBs, which remain mysterious.

The outlines of the general GRB problem have recently been reviewed in these pages⁴. Briefly, GRBs are short, intense bursts of γ -rays which occur seemingly at random across the sky. The flood of new detections by the recently launched Compton Gamma Ray Observatory has shown that their locations are consistent with an isotropic spatial distribution. Coupled with earlier indications of a relative lack of faint and hence distant GRBs (an observation confirmed by the more recent data), this result indicates that we are in the middle of a bounded distribution of bursters. This bounded distribution might represent anything from our Solar System to the Universe as a whole, but mechanisms to produce the bursts

either at the outer reaches of the Solar System or at cosmological distances are speculative at best. More plausible models, associated with surface phenomena on neutron stars, are problematic because a bounded distribution of neutron stars should be concentrated towards the centre of our Galaxy, in contradiction to the observations. As the catalogue of GRBs expands, astronomers are constantly re-evaluating the statistics of the GRB distribution, with ambiguous results⁴.

For some time, it has been recognized that three of the GRBs are different from the others. Their γ -rays are of relatively low energy (or 'soft', in the language of γ -ray astronomers) and, unlike the bulk of GRBs, they emit repeated bursts, characteristics which have led to these sources being dubbed 'soft gamma-ray repeaters' or SGRs. Radio observers then noted that the SGRs seem to be located within the remnants of supernova explosions^{5,6}, an encouraging result, as supernovae are the birthplaces of neutron stars.

Now Murakami *et al.*¹ have observed one of the SGRs with an imaging X-ray telescope, in the hope of identifying an underlying constant source of X-rays from the suspected neutron star. Not only did they succeed in this endeavour, but by good fortune they managed to catch the source in the act of bursting. These same bursts were observed by the Compton

Gamma Ray Observatory, as reported by Kouveliotou *et al.*². Thus the identity of the X-ray source as the SGR is beyond doubt. The much higher spatial resolution of the X-ray observations also leaves no doubt of the association between the SGR and the supernova remnant.

Detailed radio observations by Kulkarni *et al.*^{3,5} reveal that the supernova remnant is itself unusual. It is one of a small class of supernova remnants called 'plerions', which are illuminated by a strong central source. But this plerion also shows features reminiscent of the radio-emitting nebulae associated with several X-ray-emitting binary systems. So Kulkarni *et al.* suggest that a complete model of SGRs may need to combine elements of an unusual class of supernovae and an even more unusual class of X-ray emitters.

The circumstances leading to the creation of an SGR must be very rare. The Compton Gamma Ray Observatory has now seen bursts from all three of the previously known SGRs, but has discovered no new ones, so the census of SGRs in and near our Galaxy may be complete². The curious radio characteristics of the supernova remnant, and the high velocity of the burster implied by its off-centre location within the remnant, further suggest that something quite remarkable must be required to produce an SGR³. Although none of the three discovery teams propose a mechanism to produce the bursts, the identification of the main features of the source will provide theorists with a much firmer basis from which to explore the SGR phenomenon.

The sudden leap forward allowed by the discovery of counterparts in other energy regimes goes to show the power of 'multi-wavelength astronomy', which has so greatly aided our understanding of celestial objects ranging from clusters of galaxies to neutron stars only 10 kilometres across. The discovery of counterparts at other wavelengths remains the best hope of resolving the contentious claims surrounding the ordinary γ -ray bursts. Such an advance will be difficult: the likelihood of observing an X-ray, optical or radio counterpart of a non-repeating GRB is slim. But it is not zero, and rapid response networks are now being developed which may increase the odds. Perhaps then the bulk of these one-time flashers will be exposed for what they are, just as the softer repeat offenders have been. □

Charles Bailyn is in the Department of Astronomy, Yale University, New Haven, Connecticut 06511, USA.

1. Murakami, T. *et al. Nature* **368**, 127–129 (1994).
2. Kouveliotou, C. *et al. Nature* **368**, 125–127 (1994).
3. Kulkarni, S. R. *et al. Nature* **368**, 129–131 (1994).
4. Schaefer, B. E. *Nature* **367**, 411–412 (1994).
5. Kulkarni, S. R. & Frail, D. A. *Nature* **365**, 33–35 (1993).
6. Cline, T. L. *et al. Astrophys. J.* **255**, L45–L48 (1982).

A break in the deep

Cliff Frohlich

IN 1989 I reviewed¹ the many and various physical mechanisms that had been proposed for deep-focus earthquakes. At that time there seemed to be little hope of choosing between them. Now, only five years later, earthquake seismologists and rock mechanics experts seem to be reaching a consensus. The meeting of minds emerged at a workshop* held last month. But, while the emerging consensus represents significant progress, agreement is not yet unanimous and the deep earthquake problem cannot yet be considered solved.

The deep-earthquake issue has been one of the outstanding problems of solid-Earth geophysics for almost 70 years¹. It was in the 1920s that Kiyoo Wadati showed that some earthquakes occur hundreds of kilometres beneath the Earth's crust, while Harold Jeffreys noted that the isostatic control of topographic relief implied that the mantle had no strength at depths greater than 70 km or so. Subsequent laboratory experiments at crustal pressures and temperatures have confirmed that rocks under stress fail suddenly by brittle fracture — that is, fail as in shallow earthquakes. At higher pressures or temperatures shear stress should deform rocks by ductile flow, even in the colder regions of the mantle beneath subduction zones. So what explains

the occurrence of intermediate-focus (70–300 km) and deep-focus (300–690 km) events, which together constitute perhaps a quarter of all earthquakes?

The emerging consensus hinges on three independent assertions about the nature of intermediate- and deep-focus earthquakes. Assertion 1 is that the mechanism behind intermediate-depth earthquakes is brittle failure made possible by water subducted along with the lithosphere, or released by hydrated minerals in the crust^{2,3}. Thus, down to depths of 350 km or so, the physical mechanism is much like that of shallow earthquakes, except that the trapped fluids reduce the normal stresses on faults. They thereby allow shear failure by partially 'propping open' faults at depths unreasonable for dry fracture and friction.

Assertion 2 holds that the mechanism of failure for deep earthquakes is transformational faulting^{4–7}, in which metastable olivine transforms under stress to spinel in thin shear zones along planes of maximum shear stress. The mineral olivine, the primary constituent of the uppermost mantle, transforms under equilibrium conditions to denser spinel minerals at pressures corresponding to depths beginning about 400 km. But it is possible that olivine subducting beneath 400 km is 'supercooled' and remains as metastable olivine until it transforms catastrophically at some greater depth. There, the newly formed, fine-grained spinel is the 'fluid

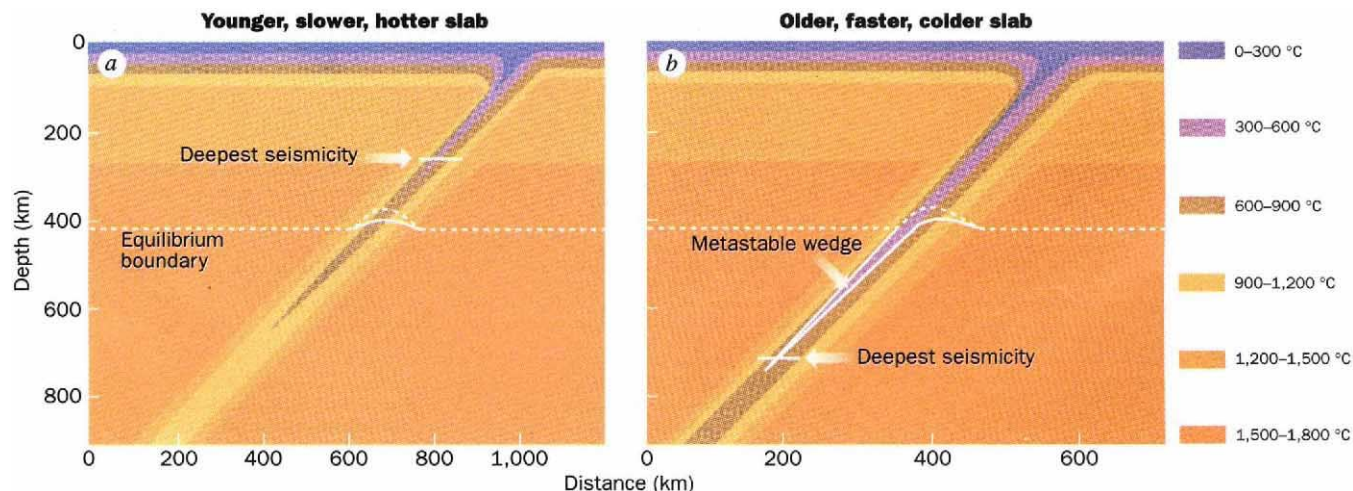
lubricant' that makes catastrophic faulting possible, the effect being augmented by heat from the exothermic olivine–spinel transformation.

Assertion 3 is that there is little or no gross difference between the source properties of deep and intermediate earthquakes, at least as determined from the analysis of their seismograms. Indeed, the only gross mechanical difference between shallow earthquakes and deeper ones is that aftershocks are much rarer for the deep and intermediate events⁸.

These three assertions are by no means universally accepted. For example, laboratory experiments⁹ have demonstrated that materials at mantle temperatures and pressures can fail in several different ways, and so it may be premature to consider any mechanism or mechanisms as established (C. Meade, Carnegie Geophysical Laboratory). Also, if two different mechanisms are responsible for deep and intermediate events, why are the respective seismograms so similar? The third assertion is troubling, then, yet perhaps it only signifies that most forms of catastrophic failure by faulting exhibit common source properties. The latest studies^{10,11} suggest that there may be at most only subtle differences in the source-time functions of deep and intermediate earthquakes (H. Houston, University of California, Santa Cruz).

Yet it is indicative of the overall consensus that various research groups are exploring the detailed implications of assertions 1 and 2. For example, thermal calculations suggest that metastable olivine persists to greater depths only in subduction zones where the slab is cold —

*Cooperative Studies of the Earth's Deep Interior Workshop: The Mechanism of Deep-Focus Earthquakes, Lake Arrowhead, California, 6–8 February 1994.



Thermokinetic model for the occurrence of metastable olivine and deep earthquakes in subducting lithosphere. *a*, Where the lithosphere is young and the subduction rate is slow (for example the Aleutians), the mantle warms the slab fast enough so that the transformation of olivine to spinel minerals keeps pace with the rate of descent and is completed near the equilibrium phase boundary at 400 km. Metastable olivine remains within a small region at most (shown by the area between the solid and dashed lines), and thus no deep-focus earthquakes occur. *b*, Older and faster-subducting lithosphere (under Tonga, for instance) remains colder, so that kinetic hindrance

prevents the olivine–spinel transformation from keeping pace with the descent rate. So metastable olivine persists in a thin wedge-shaped region well below the equilibrium phase boundary, permitting deep earthquakes to occur by transformational faulting. For the model shown, the metastable wedge is bounded on its sides and bottom by the 600 °C isotherm. The predictions of this model are consistent with the variations in maximum earthquake depth observed within individual subduction zones, and between different subduction zones. (Figure by S. Kirby, S. Stein, D. Rubie and E. Okal.)

that is, where the subducting lithosphere is old and the subduction rate is fast (see figure). This possibly explains why there are deep-focus earthquakes in Tonga and the Kuriles, whereas warmer slabs such as the Aleutians manifest intermediate-depth events only (S. Stein, Northwestern University; D. Rubie, University of Bayreuth; S. Kirby, US Geological Survey). For deep earthquakes, the most convincing support for the second assertion would be some seismological proof that metastable olivine really does persist in the mantle beneath 400 km. The presumed metastable wedge is so thin that it will be difficult to resolve it tomographically (T. Lay, University of California, Santa Cruz). It may be identifiable from analysis of waveforms¹² or travel times¹³, although the preliminary evidence is not yet definitive (J. Vidale, US Geological Survey; D. Wiens, Washington University, St Louis).

If the consensus about the mechanism for deep-focus events stands up, there is the question of whether phase transformations might provide the mechanism for intermediate-depth earthquakes as well. No, seems to be the answer, in that most specialists agree that fluid-assisted brittle fracture remains the only plausible mechanism. Although the basaltic and gabbroic minerals of the crustal part of the subducting slab, overlying the olivine-rich mantle, undergo an exothermic transformation to eclogite at appropriate depths, both theory and experiment agree that transformations of this type do not proceed quickly enough to produce catastrophic faulting. However, the accompanying density increase may provide the source of stress that causes rock failure (Kirby), just as the dehydration of crust and/or serpentinized olivine reduces normal stresses and permits fracture to occur.

Where do we go from here? No one would say that the three assertions are established unequivocally, and the sense at the workshop was that research must centre on confirming, disproving or refining them. If catastrophic transformational faulting or dehydration-assisted faulting occurs along shear zones only a few metres

thick, should we explore the possibility that relatively minor mantle constituents, other than olivine or basaltic minerals, are responsible (A. Navrotsky, Princeton University)? Is the usual seismological model that describes faulting in terms of friction- and stress-drop inappropriate for intermediate and deep earthquakes — in other words, should we interpret the data in terms of non-frictional models of failure (C. Scholz, Lamont-Doherty Geological Observatory)? Can transformational

faulting of metastable olivine somehow explain the tentative observation that the largest deep earthquakes occur preferentially at the bottoms or edges of subducting slabs, and are often distinctly isolated from neighbouring deep activity (Kirby; K. Creager, University of Washington; E. Okal, Northwestern University)? □

Cliff Frohlich is at the Institute for Geophysics, University of Texas, Austin, Texas 78759, USA.

NEUROBIOLOGY

Just a chemical attraction

Stephen J Smith

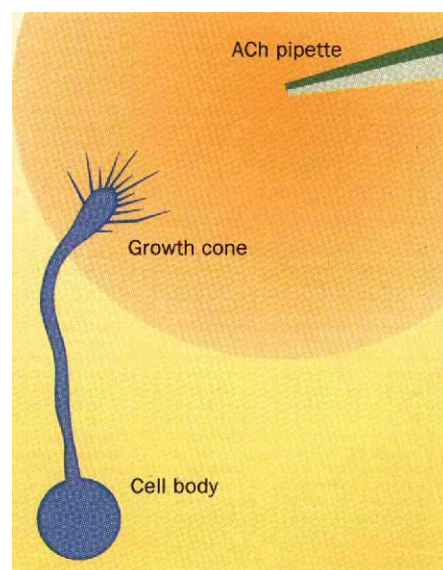
DURING development of the nervous system, axons and dendrites grow along complex but precisely defined routes. It is clear that the underlying guidance processes are multifaceted, even though our knowledge of them is still far from complete. A new facet is illuminated in the article on page 140 of this issue¹, in which Zheng *et al.* demonstrate that a neurotransmitter has chemoattractive properties: embryonic frog nerve fibres growing in cell culture are shown to follow a diffusional gradient of the neurotransmitter acetylcholine (ACh). This discovery is especially provocative because it may give us a new perspective on the mysterious ability of neuronal electrical activity to shape neural cytoarchitecture.

Cell surface and extracellular matrix molecules have been implicated as major cues in axon guidance². In addition, gradients of nerve growth factor and possibly of other diffusible factors also influence the direction of fibre growth^{3,4}. Most of these known guidance cues are probably established in an activity-independent fashion (although see ref. 5), and much of nervous-system development can indeed proceed in the absence of neural activity.

On the other hand, certain important details of neural development do depend on activity². For instance, the organization of vertebrate visual systems can be profoundly perturbed by the disruption of normal firing patterns^{2,5,6}. It has therefore been natural to ask what role activity-dependent neurotransmitter release might play in neural development. Neurotransmitters may help stabilize existing synapses (see, for example, ref. 6) and they can regulate the rate of fibre growth very powerfully⁷. Zheng and colleagues now add evidence that a neurotransmitter can regulate the direction of fibre growth. Their experiments show that a growing fibre can detect a gradient in neurotransmitter concentration, and not merely its presence.

Like many other students of nerve fibre

growth, Zheng *et al.* have looked at the growth cone. This is the highly motile expansion found at the tip of the growing axons and dendrites, and as well as being



A fibre growing from a cultured *Xenopus* neuron turns towards an acetylcholine (ACh) source (adapted from ref. 1). Filopodia fringing the growth cone may mediate response to the ACh gradient.

the main site at which fibre elongation occurs, it is centrally involved in guidance. Many embryonic guidance cues may act on processes occurring autonomously within the growth cone itself; others may initiate signals which are transmitted in some way from the cone to the cell's nucleus.

Most growth cones possess numerous highly motile filopodia. These fine, spike-like protrusions actively extend and retract from the margins of active growth cones, driven by actin-based force-generation mechanisms⁸. Filopodia consist primarily of a bundle of actin filaments covered by plasma membrane. They bear receptors⁹ and other signal-transduction machinery¹⁰, and have adhesive and

1. Frohlich, C. A. *Rev. Earth planet. Sci.* **17**, 227–254 (1989).
2. Raleigh, C. B. & Paterson, M. S. *J. geophys. Res.* **70**, 3965–3985 (1965).
3. Meade, C. & Jeanloz, R. *Science* **252**, 68–72 (1991).
4. Kirby, S. J. *geophys. Res.* **92**, 13789–13800 (1987).
5. Green, H. W. & Burnley, P. C. *Nature* **338**, 753–755 (1989).
6. Burnley, P. C., Green, H. W. & Prior, D. J. *J. geophys. Res.* **96**, 425–443 (1991).
7. Kirby, S. H., Durham, W. B. & Stern, L. A. *Science* **252**, 216–225 (1991).
8. Frohlich, C. J. *geophys. Res.* **92**, 13944–13956 (1987).
9. Meade, C. & Jeanloz, R. *Nature* **339**, 616–618 (1989).
10. Houston, H. & Williams, Q. *Nature* **352**, 520–522 (1991).
11. Vidale, J. E. & Houston, H. *Nature* **365**, 45–47 (1993).
12. Vidale, J. E., Williams, Q. & Houston, H. *Geophys. Res. Lett.* **18**, 2201–2204 (1991).
13. Iidaka, T. & Suetsugu, D. *Nature* **356**, 593–595 (1992).

tension-generating properties¹¹. Filopodia are thought to combine motor and sensory functions in the guidance of nerve growth^{9,12}. The active extension of the filopodium allows it to explore embryonic space, while its adhesion and retraction may allow it physically to steer the growth cone. Because filopodia are usually long in comparison to the dimensions of the growth cone 'body', they may provide for increased sensitivity in the detection of a spatial concentration gradient.

Zheng *et al.* used a micropipette to create well-defined ACh gradients near growth cones of fibres growing from cultured *Xenopus* spinal cord neurons (see figure). They observed the cones using a high-resolution, transmitted-light microscopy method that allowed for clear observation of the growth cone and its filopodia; they also used a fluorescence calcium-imaging method. Within two hours, most growing fibres exhibited a pronounced turning towards the ACh source.

Two much earlier effects of ACh (within minutes of gradient onset) were also evident. First, increases in the number of filopodia on the side of the growth cone facing the pipette; and, second, an increase in cytoplasmic Ca^{2+} concentration within the growth cone. In a few large growth cones, the increase in Ca^{2+} concentration sometimes appeared to be largest on the side of the cone facing the pipette. These early responses precede turning, so they may be causal. For in-

stance, ACh receptors might promote a gradient of Ca^{2+} influx parallel to the ACh gradient, promoting increased filopodial sprouting towards the pipette, and eventually leading to a net turning force. Many alternatives are possible, and experimental analysis of them should now be within reach.

Although their imaging evidence for an oriented intracellular Ca^{2+} gradient is preliminary, Zheng *et al.* buttress the idea that Ca^{2+} ions are involved in turning with complementary pharmacological experiments. These experiments implicate nicotinic ACh receptors, influx of Ca^{2+} ions across the surface membrane, and the calmodulin-dependent protein kinase II in the ACh-induced turning response. These findings are consistent with earlier indications that an experimentally created Ca^{2+} gradient within a growth cone can produce directed protrusion¹³. The molecular mechanisms by which Ca^{2+} ions might influence cytoskeletal force generation in the growth cone are just beginning to be explored¹⁴.

It is hard to speculate very seriously about the possible cytoarchitectural consequences of ACh-induced turning without knowing more about the responsive cells and fibres. Are the cells motor, sensory or interneurons? Are the fibres axons, dendrites or both? The next job will be to resolve these issues, so that the real speculation can begin.

Although the developmental significance, if any, of neurotransmitter chemo-

attraction remains to be established, a whole set of related questions is now thrust to the fore. Do significant gradients of ACh actually exist within the embryonic spinal cord or periphery? Can they be created by activity-dependent release of the neurotransmitter? Do they actually shape neuronal cytoarchitecture? If the answers turn out to be 'yes', the ability of a neurotransmitter to turn a growth cone could become central to a molecular understanding of the ways in which neural activity shapes brain structure and function. □

Stephen J Smith is in the Department of Molecular and Cellular Physiology, Beckman Center, Stanford Medical School, Stanford, California 94305, USA.

1. Zheng, J. Q., Felder, M., Connor, J. A. & Poo, M.-m. *Nature* **368**, 140–144 (1994).
2. Goodman, C. S. & Shatz, C. J. *Cell* **72**/Neuron **10** (Suppl.) 77–98 (1993).
3. Gunderson, R. W. & Barrett, J. N. *Science* **206**, 1079–1080 (1979).
4. Tessier-Lavigne, M. *Curr. Opin. Neurobiol.* **2**, 60–65 (1992).
5. Hockfield, S. & Kalb, R. G. *Curr. Opin. Neurobiol.* **3**, 87–92 (1993).
6. Cline, H. T. & Constantine-Paton, M. *Neuron* **3**, 413–426 (1989).
7. Haydon, P. G., McCobb, D. P. & Kater, S. B. *Science* **226**, 561–564 (1984).
8. Smith, S. J. *Science* **242**, 708–715 (1989).
9. Davenport, R. W., Dou, P., Rehder, V. & Kater, S. B. *Nature* **361**, 721–724 (1993).
10. Wu, D.-Y. & Goldberg, D. J. *J. Cell Biol.* **123**, 653 (1993).
11. Bray, D. *J. Cell Sci.* **37**, 391–410 (1979).
12. Bentley, D. & Toroian-Raymond, A. *Nature* **323**, 712–715 (1986).
13. Goldberg, D. J. *J. Neurosci.* **8**, 2596–2605 (1988).
14. Lankford, K. L. & Letourneau, P. C. *J. Cell Biol.* **109**, 1229–1243 (1989).

Fire, brimstone and seismicity



CAN volcanic eruptions be predicted? That age-old preoccupation is behind the work of T. P. Fischer and colleagues reported on page 135 of this issue. Their subject is the intensively monitored Galeras volcano in Colombia, the top of which is pictured above in a scene from December 1989 when the vent was virtually open and the volcano was emitting massive amounts of sulphur dioxide (some 5,000 metric tonnes a day). The authors have compared records of SO_2 flux from Galeras with those of a distinctive type of seismicity, which is thought to reflect pressure fluctuations in the plumbing system connecting the volcano's deep magma body with the surface. Galeras is

notorious for its eruption in January last year, in which several members of the monitoring teams and others died, but Fischer *et al.* concentrate on the subsequent, larger event in March. Broadly put, the observations indicate that some time before an eruption occurs an inverse correlation develops between SO_2 emission and seismicity, which can be interpreted as a gradual closing of pathways by which gases escape, and concomitant build-up of pressure inside the volcano. Such a correlation is consistent with previous evidence from other volcanoes. But the most predictable thing about volcanoes will no doubt continue to be their unpredictability. T.L.

A topological tie-in

Simon Willerton

A CENTURY ago, physicists such as Lord Kelvin¹ first delved into the theory of knots in an attempt to explain the Periodic Table, the idea at the time being that knots represented some form of stability, and that joining two knots could represent

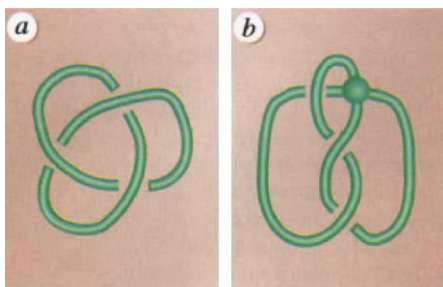


FIG. 1a, The (right) trefoil knot. This has Jones polynomial $1+x+x^3-x^4$. b, A knot with a double point or self-intersection.

the fusing of atoms. When this got nowhere, knot theory was left as an esoteric subject of study for pure mathematicians. But unexpected discoveries led to a renaissance in the 1980s; explanations for these have been trickling in from various areas, one being singularity theory, a branch of abstract mathematics, and another being quantum physics, where knots on the sub-atomic level were once again being considered. These two approaches have now been synthesized by Bar-Natan in a paper² to appear in *Topology*.

Mathematicians consider an abstraction of a knot in a piece of string that has had its ends joined to stop it being untied, so a knot is a knotted circle in three-dimensional space that does not intersect itself (Fig. 1a)³. The study of knots is part of topology, the mathematical discipline in which shapes can be deformed in any way provided that they don't tear. Two knots are said to be of the same knot type if one can be deformed into the shape of the other.

But how do mathematicians tackle such problems as "Make a list (albeit infinite) of all of the possible types of knots" and "Determine (by a finite algorithm) whether two knots are of same type"? One approach is to study knot invariants, which assign to each knot a more tractable mathematical object, say a polynomial or a number, that is the same when evalu-

ated on two knots of the same type. (A particular invariant may have the same value on two knots of different type.)

Until the mid-1980s, most knot invariants were derived from well understood topological ideas, but then an important invariant, the Jones polynomial⁴, was defined combinatorially (in other words, one draws a two-dimensional picture of the knot, then uses an algorithm to cut it into sections and thus obtain the invariant. The final step is to show that it doesn't matter what drawing was originally taken). This is a powerful invariant, in that it is good at distinguishing between knots, but at the time it lacked the topological description required if it was to be placed in a more natural context, alongside previous invariants.

Evidently, a fresh approach was needed. Vassiliev⁵, a specialist in singularity theory working in Moscow, looked at knots not as knotted circles in three-dimensional space, but as points in the abstract space of all knots. Analogously, consider a quadratic function: this can be thought of as a 'u' or 'n' shaped graph, or as being defined by the equation $f(x) = ax^2 + bx + c$, where a , b and c are real numbers. So each such function can be thought of as the point (a, b, c) in the space of all quadratics. The u-shaped graphs correspond to the triples (a, b, c) with $a > 0$ and the n-shaped graphs $a < 0$. These two regions are separated by a 'wall' — the plane $a = 0$, corresponding to functions with straight-line graphs.

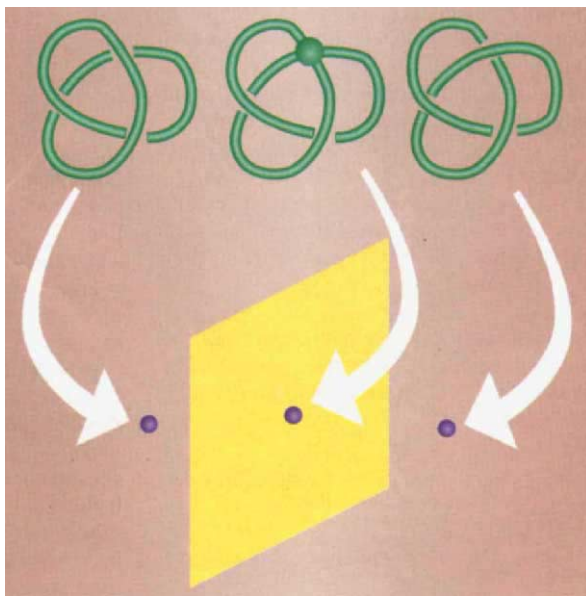


FIG. 2 A section of wall in knot space. On the right we have the trefoil knot, on the wall we have a knot with a double point and on the left a 'knot' which is unknotted. Notice how in passing through the wall we push one strand of the knot through another.

RÉSUMÉ

Seven up

THE number of examples of the earliest known bird, *Archaeopteryx*, inches towards double figures with the description of a seventh fossil (P. Wellnhofer, *Archaeopteryx* **11**, 1–30; 1993). Although all the specimens are usually thought to belong to one species, *A. lithographica*, a degree of variation hints at greater taxonomic diversity. In 1985, for example, Howgate removed the smallest ('Eichstatt') specimen to a new taxon, *Jurapteryx recurva*. Wellnhofer treads the same path in his description of the latest specimen, which yields much new information, particularly about about the jaws and sternum. The former betrays a distinctive theropod heritage, the latter the unmistakable signs of an able flyer. But more telling is its growth: although the animal was small (larger only than the Eichstatt specimen), the sternum is fully ossified, quite unlike that of the larger 'Berlin' specimen. This prompts Wellnhofer to place it in a new species, *A. bavarica*, which will make the feathers fly at the International Ornithological Congress in Vienna this August.

Storm warning

IN the search for correlations between volcanic activity and other phenomena (see opposite page), L. G. Mastin has investigated yet another possibility (*Geol. Soc. Am. Bull.* **106**, 175–185; 1994). Mastin reports that almost all of 28 'explosion-like' seismic events documented at Mount St Helens, Washington, in 1989–91 happened hours to days after storms in the area. The suggestion has been made before, but Mastin sets a figure on it; the probability that the unusually rainy weather before each eruption occurred by chance is, he says, a few per cent at most. He speculates that gas trapped beneath a shallow impermeable cap is vented after a storm because of landslides or the faster growth of cooling fractures.

Small change

S. M. BARNES and co-workers have dipped into Jim's Black Pool in Wyoming and used molecular phylogenetic techniques to open another window on biodiversity (*Proc. natn. Acad. Sci. U.S.A.* **91**, 1609–1613; 1994). The pool is a hot spring in Yellowstone National Park, and Barnes *et al.* were interested in the archaeobacteria to be found in its sediments. They extracted DNA from sediment and, using the polymerase chain reaction, ended up with ribosomal DNA sequences which they compared with sequences in a database. The number and variety of clones recovered, especially of a subgroup known as the Crenarchaeota, is remarkable — remarkable, say the authors, to the extent that the phylogenetic organization of the archaeobacteria, as understood from culture techniques, will have to be re-evaluated.

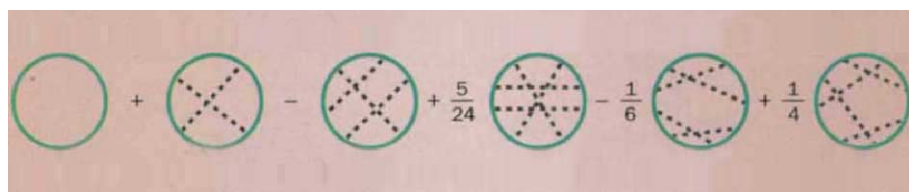


FIG. 3 The first six terms in an expansion of the trefoil knot in terms of chord diagrams. Each chord shows which parts of the circle are joined at double points (imagine shrinking the chord until the sides of the circle meet at its centre).

Similarly, a knot corresponds to a point in the space of all knots. But whereas the function space was three-dimensional in the quadratic case, the space of all knots has infinite dimensions, making it (to put it mildly) harder to visualize. In this space the different knot types are separated into chambers by walls, the walls consisting of knots that have a self-intersection or 'double point' (Figs 1b and 2 show examples). The walls meet at 'corners' which correspond to knots with two self-intersections. This analysis of walls and corners then generalizes to knots with a greater number of self-intersections.

Vassiliev approached the study of these walls by studying numerical knot invariants (invariants that assign a number to every knot, and thus a single number to each 'chamber' described above). Such an invariant can be extended to knots with a single self-intersection — in other words, its value can be defined on the walls — by looking at the knots on either side of the wall. The number on the wall is simply the difference between the values on either side. Then this procedure can be extended to knots with two double points and so on. Vassiliev invariants are those invariants that are zero when extended to knots with a certain (large) number of double points. These invariants are quite simple to understand from certain viewpoints, and Birman and Lin⁶ proved that the Jones polynomial is built up from invariants of this type.

The other approach was from quantum physics where one considers the probability of events happening: the probability of 'observables' such as the position and momentum of particles taking certain values. Around the mid-1980s physicists started considering space-times with no notion of distance. Think of two points marked on a rubber sheet: it is impossible to define a distance between the points, as stretching the sheet would alter the distance. Without the notion of distance, position and momentum become meaningless, so one needs topological ways of describing the motion of particles: in fact the only sensible thing to consider as an observable is the knottedness of the particles' trajectories.

A tool known as the Feynman path integral can be used to calculate the probabilities of particles traversing knots in three-dimensional space-time, these probabilities depending only on the type

of knot involved. A near-miraculous discovery was made by Witten⁷, a physicist at Princeton, when he was able to construct a model for space-time in which this probabilistic invariant generalized the Jones polynomial. However, although physicists happily use the Feynman integral (as it seems to give them sensible answers), it has never been properly defined mathematically. A more rigorous analysis of Witten's construction has been sorely needed.

And this is where Bar-Natan's work comes in. Working under Witten, Bar-Natan^{2,8} used perturbative methods to expand the Witten integral as a sum of terms that are easily understood mathematically, rather like expanding e^x as powers of x . This expansion is in terms of generalized Feynman diagrams (see Fig. 3); the integral describes all possible ways that the particle can behave, and each Feynman chord diagram corresponds to one of these types of behaviour. Bar-Natan has shown that the invariant derived from each chord diagram is a Vassiliev invariant, neatly uniting the two topological approaches.

Where next for the knot theorists? Kontsevich⁹ has constructed an integral that expands every knot as a sum of chord diagrams, and has proved that these methods give rise to all invariants of Vassiliev's type. His integral was originally impossible to calculate except for the simplest knots, but mathematicians are now finding ways of making it more calculable. So the latest results, although they have not answered every question, have certainly increased the arsenal with which knot theorists can attack their favourite tangles. □

Simon Willerton is in the Department of Mathematics and Statistics, Edinburgh University, Mayfield Road, Edinburgh EH9 3JZ, UK.

1. Lord Kelvin *Proc. R. Soc. Edinb.* **6**, 94–105; reprinted in *Mathematical and Physical Papers* Vol. 4, 1–12 (Cambridge Univ. Press, UK, 1910).
2. Bar-Natan, D. *Topology* (in the press).
3. Kauffman, L. H. *On Knots* (Princeton Univ. Press, 1987).
4. Jones, V. F. R. *Bull. Am. math. Soc.* **12**, 103–111 (1985).
5. Vassiliev, V. A. *Complements of Discriminants of Smooth Maps: Topology and Applications* (Trans. Math. Monogr. **98**, American Mathematical Society, Providence, 1992).
6. Birman, J. & Lin, X. S. *Invent. Math.* **111**, 225–270 (1993).
7. Witten, E. *Commun. math. Phys.* **121**, 351–399 (1989).
8. Bar-Natan, D. thesis, Princeton Univ. (1991).
9. Kontsevich, M. *Adv. Sov. Math.* **16** (2), 137–150 (1993).

DAEDALUS

Less is more

HOMOEOPATHY is a well-established fringe therapy. It offers paradoxical medicines whose effectiveness rises as their dose is reduced. The manufacturers seem curiously unconvinced by their own claim — they usually recommend a half dose for children, never the double dose required by the theory. But many formulations are so dilute that they are statistically unlikely to contain a single molecule of the active ingredient, so the precise dosage can hardly matter.

What chemical effects intensify with dilution? One candidate, says Daedalus, is electrochemical: the equilibrium potential on an electrode in solution. The Nernst equation asserts that this potential grows monotonically more negative as the concentration of the solution declines. Ultimately, zero concentration should give an infinite negative voltage. This extreme, of course, can never quite be reached. Even so, many ionic equilibria can be made so lopsided that, as in a homoeopathic remedy, there is only a small statistical chance of having even a single ion of a particular species in the solution. The Nernst negative potential is correspondingly great.

So Daedalus feels that homoeopathic remedies work, quite accidentally, by evoking Nernst potentials in the patient's tissues. They must contain impurities or trace components which release tiny concentrations of ions into the body. Common ions, such as sodium and calcium, would be swamped by the body's natural levels. Only rare and unusual ions could work: perhaps by acting on electrochemical organs such as nerves. Hence, of course, the popularity of homoeopathic treatment for the more nebulous and nervous afflictions.

DREADCO's electrochemists are now analysing homoeopathic remedies in search of traces of such non-biochemicals as dysprosium, rhenium and various exotic organometallic ions. From the results, they will design a range of rational electrochemical remedies containing traces of ions not usually present in the body, and capable of setting up high Nernst potentials across its cell walls and nerve membranes.

DREADCO's 'homoelectric remedies' should compete strongly with the traditional varieties. Like these, their high dilution will make them very economical in raw materials. But as rational medicines, they will have an intelligible biochemical action. As a result, and in strong contrast to the suspicious innocuousness claimed for so many 'natural' remedies, they may have side effects.

David Jones

Uncertainties in extinction rates

SIR — The rate at which species of plants and animals are likely to become extinct in the near future is very uncertain^{1–6}. We believe these uncertainties are compounded by difficulties in defining what is meant by 'becoming' extinct, as different authors attach quite different meanings to this deceptively simple word.

For the comparatively well-studied birds and mammals (around 13,000 species), extinction rates this century have run around one species per year; were this rate maintained, it would correspond to an average 'species lifespan' of roughly 10^4 yr. This is two to three orders of magnitude shorter than the average species lifetime of 10^6 to 10^7 yr seen over the sweep of the fossil record⁷, but it is almost two orders of magnitude longer than the impending extinction times — characteristically 200–400 yr — which have been deduced in three independent ways: (1) from observed rates of habitat loss combined with heuristic species–area relations^{4,8,9}; (2) by extrapolating the rates at which species appear to be 'climbing the ladder' of IUCN-defined categories of threat¹⁰ from 'vulnerable' to 'endangered' to 'extinct'²; and (3) by using the assessments of species-by-species extinction probability distributions, as functions of time, which underly the new draft IUCN criteria^{11,12} for categories of threat, to calculate the expected times when half the species in each of 10 vertebrate taxa will be extinct (three, four, three orders or families of reptiles, birds, mammals, respectively; these 10 average extinction times range from 100 to 1,000 yr, but are characteristically 300–400 yr for mammals and birds)^{10,12}. All three of these theoretical estimates are admittedly shaky: it is doubtful to what extent the effects of habitat fragmentation, especially in the tropics, are described by species–area relations based on empirical observations of species distributions among island archipelagoes^{9,13}; and estimates based on IUCN 'category-ladder-climbing' or extinction probability distributions may tell us much more about patterns of human concern, the taxonomic group in question,

and rates of data entry than about extinction rates^{2,3}. Even so, the discrepancy between three different but concordant theoretical estimates of impending extinction rates, and the rates currently documented, is disturbing.

This discrepancy largely disappears if we are more careful in recognising that the species–area or 'ladder-climbing' estimates are projected numbers of species that, on current trends, are 'committed' to extinction. In particular, species–area relations suggest that destroying 90% of a habitat will eventually lead to loss of around half the species in it, but the time taken to attain this new, impoverished state is not specified, and will depend on details that vary from place to place and species to species. In general, the notion of 'committed' to extinction recognizes that extinction is a gradual process on an uncertain timescale, but that the end point is nevertheless certain once a species' demographic and genetic base has been sufficiently eroded by habitat destruction or other processes.

Simberloff¹⁴, for example, used a species–area approach to estimate that about 1,350 species of birds would be committed to extinction by the year 2015. If this figure were misinterpreted as the predicted number actually to become extinct between 1986 and 2015 (which was not Simberloff's intent), it would imply an average of 45 extinctions per year, which seems implausibly high. A more empirical approach is to examine the 1,029 entries in Collar and Andrews¹⁵ list of threatened bird species. Interpreting 'committed' to extinction as referring to any species whose populations in the wild are no longer viable and will inevitably become extinct, unless major conservation actions reverse current trends (by habitat restoration, elimination of introduced predators, captive breeding and re-introduction, and so on), Two of us estimated¹³ that 450 bird species will be committed to extinction by 2015, with 27 of these already probably extinct (in addition to the 12 listed as almost certainly extinct¹⁵). Although this empirically based estimate is still one-

third the theoretically based one, the disparity is not the worrying two orders of magnitude noted earlier.

In short, projected extinction rates are beset by many uncertainties. But theoretical estimates and species-by-species studies for birds give answers that are roughly concordant, provided we are clear about the stages involved in 'becoming' extinct. Broadly similar answers pertain to the insect fauna of the United Kingdom^{16,17}, although in general our knowledge of the endangered status of invertebrates is woefully inadequate¹⁸. The situation is less clear for plant species, and seems to depend a lot on the area in question. Moreover, the genetic erosion of plant species which are not yet committed to extinction is a major concern, yet is generally overlooked in the extinction debate.

Vernon H. Heywood,
BGCI, Descanso House,
199 Kew Road, Richmond,
Surrey TW9 3BW, UK
Georgina M. Mace
Institute of Zoology,
Regent's Park, London NW1 4RY, UK
Robert M. May
Department of Zoology,
University of Oxford,
Oxford OX1 3PS, UK
S. N. Stuart
Species Survival Commission,
IUCN, Gland, Switzerland CH-1196

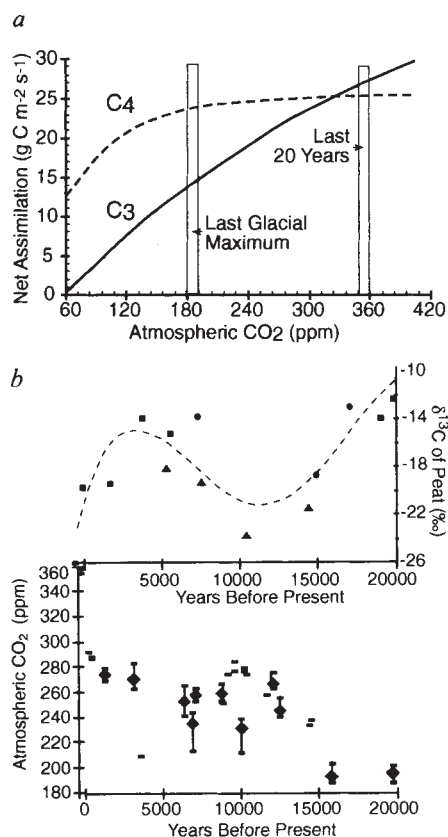
Atmospheric CO₂ and plants

SIR — Under low atmospheric-CO₂ concentrations ($p\text{CO}_2$), C4 plants have a physiological advantage over C3 plants¹. Collatz *et al.* have published models^{2,3} appropriate for predicting the magnitude of this advantage. These (*a* in the figure) give C4 plants a major advantage at $p\text{CO}_2$ values characteristic of the Pleistocene glaciations (190–240 p.p.m.; ref. 4). Experimental work confirms that $p\text{CO}_2$ of the order of 200 p.p.m. is stressful to C3 plants, reviewed in ref. 5. Hence, all other things being equal, one would expect C4 niche expansion during the ice ages. C4 physiology, however, is inefficient at low temperatures^{1,2}, and in many regions, shrinkage of the C4 niche due to ice-age cooling probably outweighed expansion due to CO₂ starvation.

The claim that low CO₂ promotes C4 expansion at high altitude is equivocal¹. Historical $\delta^{13}\text{C}$ data may offer a stronger line of evidence. Data from ref. 6 on $\delta^{13}\text{C}$ for a 20,000-year southern Indian peat core are juxtaposed to $p\text{CO}_2$ history as reconstructed from ice cores^{4,7} (*b* in the figure). As shown, and as would be anticipated if C4 plants displaced isotopically lighter C3 competitors under CO₂ starvation, $\delta^{13}\text{C}$ and $p\text{CO}_2$ are inversely related.

- Whitmore, T. C. & Sayer, J. A. (eds) *Tropical Deforestation and Species Extinction* (Chapman and Hall, London, 1992).
- Smith, F. D. M. *et al.* *Nature* **364**, 494–496 (1993).
- Cauron, A. D. *Nature* **366**, 118 (1993).
- Wilson, E. O. *The Diversity of Life* (Harvard University Press, 1991).
- Diamond, J. M. *Phil. Trans. R. Soc. B* **325**, 469–477 (1989).
- Reid, W. R. & Miller, K. R. *Keeping Options Alive: The Scientific Basis for Conserving Biodiversity* (World Resources Institution, Washington, DC, 1989).
- Raup, D. M. *Paleobiology* **4**, 1–5 (1978).
- Reid, W. V. in *Tropical Deforestation and Species Extinction* (eds Whitmore, T. C. & Sayer, J. A.) 55–73 (Chapman and Hall, London, 1992).
- Simberloff, D. in *Tropical Deforestation and Species Extinction* (eds Whitmore, T. C. & Sayer, J. A.) 75–89 (Chapman and Hall, London, 1992).

- Mace, G. in *Large Scale Ecology and Conservation Biology* (eds P. H. Edwards *et al.*) 295–314 (Blackwell, Oxford, 1994).
- Seal, U. S., Foose, T. J. & Ellis-Joseph, S. in *Creative Conservation* (eds P. J. Olney *et al.*) 312–325 (Chapman and Hall, London, 1993).
- Mace, G. M. *Phil. Trans. R. Soc.* (in the press).
- Heywood, V. H. & Stuart, S. N. in *Tropical Deforestation and Species Extinction* (eds Whitmore, T. C. & Sayer, J. A.) 91–117 (Chapman and Hall, London, 1992).
- Simberloff, D. in *Dynamics of Extinction* (eds Elliott, D. K.) 165–180 (Wiley, New York, 1986).
- Collar, N. J. & Andrew, P. *The ICBP Checklist of Threatened Bird Species* Tech. Pub. No. 8 (ICBP, Cambridge, 1988).
- May, R. M. in *Action for Biodiversity in the UK* (eds Wright, F. J. *et al.*) 16–20 (JNCC, Peterborough, 1993).
- Shirt, D. B. *British Red Data Book No. 2: Insects* (JNCC, Peterborough, 1987).
- Stork, N. E. & Lyl, C. H. C. *Nature* **366**, 307 (1993).



a, Predicted net photosynthesis assuming 25°C air temperature, 2.5 kPa ambient partial pressure of water vapor, 10 mol m⁻² s⁻¹ boundary layer conductance, C3 and C4 V_{max} for rubisco, respectively, of 100 and 30 μ mol m⁻² s⁻¹, and 1.64 mmol m⁻² s⁻¹ photon flux density (~80% of full sunlight). Calculated by G. J. Collatz based on models documented in refs 2 and 3. **b**, Atmospheric CO₂ concentrations from two Antarctic ice cores as related to $\delta^{13}\text{C}$ measurements from a southern Indian peat core⁶. CO₂ data without range bars from ref. 4 and with range bars from ref. 7.

Remarkably, the data show near-pure C4 peat (~12.8‰) at the last glacial maximum (18,000 years before present).

By extension, the combination of low $p\text{CO}_2$ with hot, sunny growing seasons should yield high- $\delta^{13}\text{C}$ peat; the implications of other atmospheric conditions can be predicted with photosynthetic models. Analogous reasoning, using lignites in place of peats, might add pieces to the puzzle of Miocene C4 expansion^{8,9}.

Jennifer M. Robinson

UFZ-Centre for Environmental Research,
Leipzig-Halle GmbH, 04318 Leipzig,
Germany

1. Moore, P. D. *Nature* **367**, 322 (1994).
2. Collatz, G. J., Ribas-Carbo, M. & Berry, J. A. *Aust. J. Plant Physiol.* **19**, 519–538 (1992).
3. Collatz, G. J., Grivet, G., Ball, J. T. & Berry, J. A. *Ag. Forest Meteorol.* **54**, 107–136 (1991).
4. Barnola, J. M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. *Nature* **329**, 408–414 (1987).
5. Robinson, J. M. *Pl. Cell Envir.* **17**, 1–10 (1994).
6. Sukumar, R. *et al.* *Nature* **364**, 703–705 (1993).
7. Leuenberger, M., Siegenthaler, U. & Langway, C. C. *Nature* **357**, 488–490 (1992).
8. Morgan, M. E. *et al.* *Nature* **367**, 162–164 (1994).
9. Cerling, T. E. *et al.* *Nature* **361**, 344–345 (1993).

Similarity of *met* and *trp* repressors

SIR — Lawson and Carey¹ describe the three-dimensional structure of a complex between *Escherichia coli trp* repressor and its operator, where repressors are bound in a tandem array to the DNA. The form of this array, and its functional consequences, are directly analogous to those observed for the *met* repressor system, as we have predicted^{2,3}. We have also reported the three-dimensional structure of a tandem *met* repressor-operator complex⁴, and have confirmed its relevance to repression *in vivo*⁵.

Both *met* and *trp* repressors recognize unusual operator regions, with tandem repeats of imperfect 8-base-pair (bp) boxes. The 'met box' has consensus sequence AGACGTCT. This is an inverted repeat sequence, that is, it has an axis of 2-fold symmetry at its centre between bases C and G. Tandem strings of met boxes additionally have 2-fold axes at the junctions between boxes. This symmetry could exactly match that of a superhelical cooperative array of tandem repressors wound around the DNA, each repressor being a 2-fold symmetric protein dimer^{2,3}. This arrangement of 2-fold axes alternating at 4-bp intervals leads to an ambiguity in assignment of the correct repressor site^{2,3}. There is a direct similarity to *trp* repressor, which binds to arrays of 'trp boxes' (consensus AGTTAACT) with identical symmetry properties^{6,7}. Staacke *et al.*⁸ subsequently examined this ambiguity for *trp* repressor, but did not comment on the parallels with *met* repressor.

We also noted that consensus *trp* and met boxes share 50% sequence identity. The two repressors are similar in controlling amino-acid biosynthesis and binding small corepressor molecules. Intriguingly, the repressors themselves have totally different polypeptide folds, and contact DNA through different secondary structure motifs, α -helices in *trp* repressor, and β -strands in *met* repressor.

In our crystal structure for the tandem *met* repressor-operator complex⁴ (atomic coordinates available from the Brookhaven Protein Data Bank as entry 1cma), the geometric parameters describing the repressor superhelical array, and the helical order of the DNA, are almost identical to those now reported for *trp* repressor.

Lawson and Cary also raise the question of whether such tandem binding is important *in vivo*. We have shown that the cooperativity generated by the formation of the repressor array is essential to effective repression in the *met* case by changing amino-acid residues in the interface between adjacent repressors in the array

using site-directed mutagenesis, and demonstrating severe impairment of repression efficiency in an assay *in vivo* using a reporter gene. We would expect similar effects with some *trp* operators. In fact, the cooperativity generated by the tandem binding produces a very sensitive genetic switch, which may be thrown by small variations in corepressor concentration. We can expect to find further examples among gene regulatory proteins, and there are parallels with the assembly of the multiprotein complexes on the DNA before initiation of transcription by RNA polymerase.

Simon E. V. Phillips

Department of Biochemistry & Molecular Biology

Peter G. Stockley

Department of Genetics

University of Leeds,

Leeds LS2 9JT,

UK

1. Lawson, C. L. & Cary, J. *Nature* **366**, 178–182 (1993).
2. Phillips, S. E. V. *et al.* *Nature* **341**, 711–715 (1989).
3. Rafferty, J. B., Somers, W. S., Saint-Girons, I. & Phillips, S. E. V. *Nature* **342**, 705–710 (1989).
4. Somers, W. S. & Phillips, S. E. V. *Nature* **359**, 387–393 (1992).
5. He, Y.-Y. *et al.* *Nature* **359**, 431–433 (1992).
6. Kumamoto, A. A., Miller, W. G. & Gunsalus, R. P. *Genes Dev.* **1**, 556–564 (1987).
7. Bass, S., Sugiono, P., Arvidson, D. N., Gunsalus, R. P. & Youderian, P. *Genes Dev.* **1**, 565–572 (1987).
8. Staacke, D., Walter, B., Kisters-Wolke, B., Wilcken-Bergmann, B. V. & Müller-Hill, B. *EMBO J.* **9**, 1693–1697 (1990).

Future prospects discussed

SIR — "There are lies, damn lies and statistics" is one of those colourful phrases that bedevil poor workaday statisticians who labour believing that they contribute to the advancement of scientific knowledge. In my view, the statistical methodology of Gott¹, which purportedly enables one to put statistical limits on the probable lifetime of everything from human existence to *Nature* itself, breathes unfortunate new life into the saying.

Gott claims that, given the duration of existence of anything, there is a 5% probability that it is in its first or last 2.5% of existence. He uses this logic to predict, for example, the duration of publication of *Nature*. Given that *Nature* had published for 123 years when he wrote his article, he projects the duration of continued publication to be between 123/39=3.2 years and 123 × 39=4,800 years, with 95% certainty. He then goes on to predict the future longevity of our species (5,000–7.8 million years), the probability we will colonize the Galaxy and the future prospects of space travel.

This technique would be a wonderful contribution were it not based on a fallacious argument. Dubbed the "Principle

of Indifference" by John Maynard Keynes in the 1920s, and the "Principle of Insufficient Reason" by Laplace in the early 1800s, it has its origins as far back as Leibniz in the 1600s (ref. 2). Among other counterintuitive results, this principle can be used to justify the prediction that after flipping a coin and finding a head, the probability of a head on the next toss is $2/3$ (ref. 3). It has been the source of many an apparent paradox and controversy, as alluded to by Keynes, "No other formula in the alchemy of logic has exerted more astonishing powers. For it has established the existence of God from total ignorance, and it has measured with numerical precision the probability that the sun will rise tomorrow"⁴. Perhaps more to the point, Kyburg, a philosopher of statistical inference, has been quoted as describing it as "the most notorious principle in the whole history of probability theory"⁵.

Simply put, the principle of indifference says that if you know nothing about a specified number of possible outcomes, you can assign them equal probability. This is exactly what Gott does when he assigns a probability of 2.5% to each of the 40 segments of a hypothetical lifetime. There are many problems with this seductively simple logic. The most fundamental one is that, as Keynes said, this procedure creates knowledge (specific probability statements) out of complete ignorance. The practical problem is that when applied in the problems that Gott addresses, it can justify virtually any answer. Take the *Nature* projection. If we are completely uncertain about the future length of publication, T , then we are equally uncertain about the cube of that duration T^3 . Using Gott's logic, we can predict the 95% probability interval for T^3 as $T^3/39$ to $39T^3$. But that translates into a 95% probability interval for the future length of publication to be $T/3.4$ to $3.4T$ ($3.4 = \sqrt[3]{39}$), or 42 to 483 years, not 3 to 4,800. By increasing the exponent, we can conclude that we are 95% sure that the future length of anything will be exactly equal to the duration of its past existence, T . Similarly, if we are ignorant about successively increasing roots of T , we can conclude that we are 95% sure that the future duration of anything will be somewhere between zero and infinity. These kinds of problems are inherent in any argument based on the principle of indifference.

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references.

On the positive side, we should all be encouraged to learn that there can be no meaningful conclusions where there is no information, and that the labours of scientists to predict such things as the survival of the human species cannot be supplanted by statistical arguments.

Steven N. Goodman
Oncology Center,
Division of Biostatistics,
Johns Hopkins School of Medicine,
Baltimore, Maryland 21205, USA

1. Gott, J. D. III *Nature* **363**, 315–319 (1993).
2. Hacking, I. *The Emergence of Probability* p. 126 (Cambridge Univ. Press, 1975).
3. Howson, C. & Urbach, P. *Scientific Reasoning: The Bayesian Approach*, 100 (Open Court, La Salle, Illinois, 1989).
4. Keynes, J. M. *A Treatise on Probability* p. 89 (Macmillan, London, 1921).
5. Oakes, M. *Statistical Inference: A Commentary for the Social Sciences*, 40 (Wiley, New York, 1986).

SIR — At a lecture in Cambridge, Steven Wolfram was asked whether he could supply a meter which would indicate how near a Mathematica computer program was to finishing. Wolfram explained that there were serious philosophical problems preventing this very desirable facility. Now J. R. Gott III (*Nature* **363**, 315–319; 1993) has used the Copernican principle (that we do not occupy a privileged position in the Universe) to estimate when human life might come to an end. Bearing in mind E. Fredkin's view that the Universe is itself a computer, Gott's method can be applied directly to the computer case.

We start an unfamiliar computer program and it has not halted after one unit of time (minute, hour or day according to our patience). On the assumption that *time (now)* is uniformly distributed in the range *time (begin)* to *time (end)*, then at the 50% confidence level, Gott finds that *time (future)* < 3 *time (past)*, or at the 95% confidence level, *time (future)* < 39 *time (past)*. This gives a good working rule. If your program has not finished in 1 hour there is an even chance that it will still not have finished in three further hours.

Alan L. Mackay
Department of Crystallography,
Birkbeck College,
University of London,
Malet Street,
London WC1E 7HX

SIR — In his Hypothesis (*Nature* **363**, 315–319; 1993), J. R. Gott III defines an intelligent observer as a being capable of understanding and applying his "delta t argument". In this sense I am not such a being. I think the delta t argument is wrong.

Gott claims to have derived some equations that allow him to calculate an estimate for the future longevity of any system just from the knowledge of its past duration. He applies them, with hind-

sight, to Stonehenge, the Berlin Wall and the Soviet Union. In these selected examples his theory turns out to be correct. From a present point of view, he predicts the future publication of *Nature* and the longevity of mankind. His results were surprising, almost unbelievable.

Let us rewrite Gott's argument in the strict terms of probability theory. Consider that the observation of a system is possible in a time interval of length T . We do not know T , and we can only speculate about its probability distribution $P(T)$. The time of observation divides the interval into two parts of length T_{past} and T_{future} , respectively. It is random in the sense that (for a given T) the values for T_{past} are equally distributed in the interval $[0, T]$. That is,

$$P(T_{\text{past}}=t_p|T=t) dt_p = \begin{cases} (1/t) dt_p, & 0 < t_p < t \\ 0 & \text{else} \end{cases} \quad (1)$$

Here, the left-hand side of equation (1) is the symbolic notation of the condition probability distribution of the variable T_{past} for T having the value t .

For the future-to-past ratio R , $R = T_{\text{future}}/T_{\text{past}}$, we find

$$P(R=r|T=t) dr = P(T_{\text{past}}=t_p|T=t) |dt_p/dr| dr = (r+1)^{-2} dr, r>0 \quad (2)$$

and

$$P(R>r|T=t) = \int_r^\infty P(R=r'|T=t) dr' = (r+1)^{-1}, r>0 \quad (3)$$

This is the result of Gott's equation (6). In this notation, Gott's equation (1) reads

$$P(1/39 < R < 39|T=t) = (1/39+1)^{-1} - (39+1)^{-1} = 0.95 \quad (4)$$

This equation means that for a given total time T (in a sample of systems with an arbitrary, but common and fixed T) the future duration of a system is between $1/39$ and 39 times its past duration in 95% of all cases, which is obviously true. Although the right-hand sides of equations (2)–(4) do not depend on t , there is no way to omit the condition $|T=t$ on the left-hand sides.

Alas, the equations above do not apply to the intelligent observer's situation. He does not know T , but he knows T_{past} . So we have to calculate

$$P(R=r|T_{\text{past}}=t_p) = \int_0^\infty P(R=r|T=t \text{ AND } T_{\text{past}}=t_p) P(T=t|T_{\text{past}}=t_p) dt \quad (5)$$

The additional condition AND $T_{\text{past}}=t_p$ must not be omitted.

Using

$$P(R=r|T=t \text{ AND } T_{\text{past}}=t_p) = \delta(r-(t-t_p)/t_p)$$

and Bayes' theorem

$$= (\int P(a|b')P(b')db')^{-1} P(a|b) P(b) db$$

for the second factor in the integral, we find

$$P(R=r|T_{\text{past}}=t_p)dr = (\int_0^\infty P(T=t)t^{-1}dt)^{-1} P(T=(r+1)t_p)(r+1)^{-1} dr \quad (6)$$

Contrary to equation (2), this depends on the *a priori* distribution $P(T)$ (this is the deeper meaning of Bayes' theorem: you cannot invert a condition probability distribution unless you have an *a priori* probability). It can easily be proved that there is no normalized probability distribution that makes equations (2) and (6) match.

Now consider, as a special case, a constant extinction rate λ_0 . The distribution of intervals of potential observation is exponential, and for an actually observed interval

$$P(T=t)dt = \lambda_0^2 t \exp(-\lambda_0 t) dt \quad (7)$$

where an 'anthropic' factor of t accounts for the higher probability of observing a longer interval. Inserting equation (7) into equation (6) yields

$$P(R=r|T_{\text{past}}=t_p)dr = \lambda_0 t_p \exp(-\lambda_0 t) dr \quad (8)$$

For the variable T_{future} ,

$$P(T_{\text{future}}=t_f|T_{\text{past}}=t_p)dt_f = \lambda_0 \exp(-\lambda_0 t_f) dt_f \quad (9)$$

Hence the future behaviour of a system with exponential decay does not depend on its past, but only on its intrinsic parameter λ_0 . This result is well known from particle physics.

To summarize, Gott first treated the past and future of a system in a symmetric manner. Later, he switched over to a derivation of the future from the past. Thus he changed the conditions under which his equations were derived. By keeping his equations unchanged he obtained some surprising results. Here I have tried to show that these results are wrong. Gott's prediction of the future from nothing but the past does not reflect the reality we are all living in, including intelligent observers.

P. Buch

Friedrichstraße 23 b,
D 53332 Bornheim, Germany

GOTT REPLIES — Goodman fails to note that while Keynes did show some examples where he thought it was inappropriate to apply the principle of indifference, he also gave examples where even he thought the principle of indifference could be used safely. He said¹: "Suppose, for instance, that a point lies on a line of length $m.l.$, we may write the alternative 'the interval of

length l on which the point lies is the x th interval of that length as we move along the line from left to right' = $\Phi(x)$; and the Principle of Indifference can then be applied safely to the m alternatives $\Phi(1), \Phi(2), \dots, \Phi(m)$, the number m increasing as the length of l of the intervals is diminished. There is no reason why l should not be of any definite length however small." My application², which involves dividing the observed interval into 40 equivalent subintervals of equal length, is similarly conservative. Statisticians have long argued about which axioms to adopt in this context. According to the Copernican principle, you expect your observation of whatever you are measuring to be located randomly on the chronological list of such observations.

Buch would like to know the expected (posterior) probability distribution for the total longevity L of our species after we observe that our past age is $t_p = 200,000$ yr, using Bayesian statistics³.

$$P_{\text{posterior}}(L) dL \propto \text{Likelihood} (t_p = 200,000 \text{ yr}) \times P_{\text{prior}}(L) dL \quad (1)$$

Often P_{prior} is based on some real data, but as we have no data on the longevities of other intelligent species we are not free to adopt just any prior but are commanded to adopt an appropriate vague Bayesian prior. We know that L is positive but we do not know anything about its magnitude. In such a situation we should adopt as the appropriate vague Bayesian prior³

$$P_{\text{prior}}(L)dL \propto d\ln L \propto L^{-1} dL \quad (2)$$

This means that there is a *priori* an equal chance of L being in any order of magnitude interval. The likelihood of observing $t_p = 200,000$ yr is 0 if $L < 200,000$ years, so

$$P_{\text{posterior}}(L)dL = 0 \text{ if } L < 200,000 \text{ yr} \quad (3)$$

And the likelihood of observing $t_p = 200,000$ yr if $L > 200,000$ yr is proportional to L^{-1} , so

$$P_{\text{posterior}}(L)dL \propto L^{-1} \times L^{-1} dL \propto L^{-2} dL \text{ if } L > 200,000 \text{ yr} \quad (4)$$

Pick a longevity $L_0 > 200,000$ yr. Integrating $P_{\text{posterior}}(L)dL$ from L_0 to infinity and normalizing, we find that

$$P_{\text{posterior}}(L > L_0) = (200,000 \text{ yr}/L_0) \text{ for } L_0 > 200,000 \text{ yr} \quad (5)$$

Thus, $P_{\text{posterior}}(L > 200,000 \text{ yr}) = 1$ and $P_{\text{posterior}}(L > 400,000 \text{ yr}) = 0.5$, and so forth. Define Y by the equation $L_0 = (Y + 1)t_p = (Y + 1) \times 200,000$ yr. Because $L = t_f + t_p = [(t_f/t_p) + 1]t_p$, $L > L_0$ is equivalent to $(t_f/t_p) > Y$ so

$$P_{\text{posterior}}([(t_f/t_p) > Y]) = 1/(Y + 1) \quad (6)$$

which is identical to the distribution found

in the delta t argument². It is not necessary that the vague Bayesian prior be normalizable, just that the posterior probability is³. If one still demanded a normalized prior one could establish lower and upper cutoffs of 10 yr and $10^{5,000,000}$ yr, respectively, corresponding to the longevity of an individual and of our inflationary domain in the Linde cosmology, without significantly altering the results.

Buch's second question concerns the problem of a constant extinction rate λ_0 . He is right that if one knew λ_0 , one would have the perfect prior and would be able to determine the probability distribution for t_{future} $P(t_f)dt_f = \lambda_0 \exp(-\lambda_0 t_f) dt_f$, independently of the particular observed value of t_{past} in this case. But we do not know λ_0 . We may, however, learn something about it by inspecting t_{past} . Because the dispersions in the relevant galactic and stellar evolution timescales are measured in billions of years and we originated approximately midway in the main sequence lifetime of the Sun, the rate at which intelligent species have been forming in the Universe over the past millions of years has been approximately constant⁴. Thus, $P(t_p) dt_p = \lambda_0 \exp(-\lambda_0 t_p) dt_p$, because, of all those intelligent species formed at an epoch t_p ago, only a fraction $f = \exp(-\lambda_0 t_p)$ are still around today². Thus $\langle t_p \rangle = \lambda_0^{-1}$ and there is a 95% chance that $\lambda_0^{-1} < 19.5 t_p$. Also $P(t_p, t_f) dt_p dt_f = \lambda_0^2 \exp(-\lambda_0 [t_p + t_f]) dt_p dt_f$, giving $P(L)dL = \lambda_0^2 L \exp(-\lambda_0 L) dL$ (where $L = t_p + t_f$) and the distribution in eq. (6).

Mackay misquotes the delta t argument. The correct statement is that there is a 50% chance that the future longevity lies between $1/3$ and 3 times the past longevity and a 95% chance that the future longevity lies between $1/39$ and 39 times the past longevity as Buch correctly noted in his equation. Because Mackay started the program and picked the time interval after which to look, it could be argued that it might be special because he may have become impatient only when it had run longer than others, or he may have consciously picked a wait time shorter than most programs. Here's an example that should work, however. If Mackay happened to have a program running when he first heard of my paper, then the delta t argument should tell him how much longer after that the program would run: if it had been running for 1 hour, he could say that there was a 50% chance that it would continue to run for more than 20 minutes but less than 3 hours.

J. Richard Gott, III

Department of Astrophysical Sciences,
Princeton University,
Princeton, New Jersey 08544, USA

1. Keynes, J. M. in *A Treatise on Probability*, P. 62 (AMS, New York 1979).
2. Gott, J. R. III *Nature* **363**, 315–319, (1993).
3. Press, S. J. *Bayesian Statistics* 25–52 (Wiley, New York, 1989).
4. Carter, B. *Phil. Trans. R. Soc. A* **310**, 347–363 (1983).

Intellectual mastery over nature

Peter Harman

Hermann von Helmholtz and the Foundations of Nineteenth-Century Science. Edited by David Cahan. *University of California Press: 1994. Pp. 666. \$65.*

HERMANN von Helmholtz (1821–94) was a scientific polymath of extraordinary accomplishment and achievement, the dominant figure of German science in the second half of the nineteenth century. His contributions ranged over physics and physiology, from the first definitive statement of the law of conservation of energy to work in physiological optics (which included the invention of the ophthalmoscope), physiological acoustics (and its bearing on the theory of music), colour theory and colour vision. First trained in medicine, he nevertheless managed in his work to embrace mathematical physics, producing famous papers on hydrodynamics, electromagnetism and thermodynamics; and he made important contributions to non-Euclidean geometry and the philosophy of science.

Helmholtz stood at the pinnacle of technical accomplishment: he designed new instruments, carried out precise experiments and employed abstract mathematical analysis. Further, he stood out as the leading spokesman for German science, contributing substantially to the popularization of science and emphasizing its civilizing power. As David Cahan aptly remarks, Helmholtz represented the values of the Enlightenment and the liberal tradition in German culture. He also spoke for the conservative academic and cultural élite, explaining the role of science within traditional culture and its centrality in the emergent industrial economy. He was also a truly international figure. By the early 1850s he had made contact with British scientists, forming a friendship with the young William Thomson (later Lord Kelvin). Members of Thomson's circle, especially the xenophobic Peter Guthrie Tait, were not always sympathetic to the endeavours of German colleagues; but Helmholtz was respected, even revered.

In an autobiographical sketch written towards the end of his life, Helmholtz recounted what had first drawn him to science. This was the desire to attain "intellectual mastery of nature", a control to be obtained through understanding the laws of nature. Helmholtz's abiding intellectual interests crossed scientific disciplines, and his work is characterized by his ability to unify a field of investigation and to link ideas between different sciences. Broad unifying themes engaged his attention: human perception and cognition, energy transformations and the concept of nature as a mechanical system. He

therefore fastened onto the discussion of scientific problems at the boundaries of scientific disciplines. And he aimed to set scientific concepts within a broader framework of philosophical investigation that reached out to embrace aesthetic experience and express the cultural values of scientific inquiry.

Given Helmholtz's supreme importance in the history of science and culture, it is surprising that he has received relatively little attention from historians. This is no doubt due to the awesome

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Hermann von Helmholtz (1821–94)

breadth of his scientific work and intellectual endeavour. To have such a range in any century would be remarkable; but to attain mastery and intellectual leadership in such diverse scientific disciplines in the nineteenth century is wholly astonishing. For in this period science truly came of age as a professional and specialized activity. In a scientific portrait published in *Nature* in 1877, James Clerk Maxwell described Helmholtz as an "intellectual giant", and observed that few scientific specialists could reach the level of attainment he had achieved in any of his diverse papers. Maxwell pointed to Helmholtz's "thoroughness which of itself demands the mastery of many sciences, and in so doing makes its mark on each". For any individual historian to achieve a level of thoroughness and mastery appropriate to this subject is no easy task, and the present collection of 15 essays by historical spe-

cialists is a valid and successful response to the demand for a serious evaluation of Helmholtz's scientific achievement.

The contributors are historians of science from Europe and North America, and include leading scholars in the field such as Frederic L. Holmes and Jed Buchwald, as well as some younger historians. The collection is divided into three parts, on Helmholtz as physiologist, physicist and philosopher. In keeping with the crossdisciplinary style of Helmholtz's own work, many of the essays necessarily transcend narrow scientific specialities, and touch on problems in his work on other topics, and on the overarching philosophical themes that characterize his approach to the intellectual mastery of nature.

Several of the essays indicate that Helmholtz paid a price for the broad range of his scientific research. His leadership in a given discipline often proved to be short-lived. In the first half of his career he was professionally a physiologist, but he moved to Berlin in 1871 to assume a leading position in the physics community, largely abandoning work in physiology and medicine. In the study of colour vision his work of the early 1850s was in part transcended by Maxwell, and he essentially abandoned research on the subject after 1860. He had already made important contributions to physics, with his famous essay on energy conservation of 1847 and his influential mathematical study of vortex motion of 1858, but he had been content to leave the development of energy physics and hydrodynamics to others. His assessment in a key paper of 1870 of the merits of Maxwell's field theory did, however, shape his direction of research on electrodynamics in his Berlin laboratory, an interest that ultimately bore fruit in Heinrich Hertz's famous experiments that provided evidence for Maxwell's electromagnetic theory of light.

But in neither this work nor his work on thermodynamics did Helmholtz maintain his own interest or sustain his leadership. If his main scientific achievements belong to the first half of his career, it is worth remembering that on his move to Berlin he was 50 years old, and had published his principal treatises on physiological optics and physiological acoustics. In physics he achieved no similar embracing systematic statements of theory and practice. His greatness lies in the range and sophistication of his science, in his successful quest for intellectual synthesis and in the technical accomplishment that characterizes his scientific work. His evocation of the civilizing power of science and of the values of intellectual freedom continue to speak to us. □

Peter Harman is in the Department of History, University of Lancaster, Lancaster LA1 4YG, UK.

Head bones

R. McNeill Alexander

The Skull. Volumes 1–3. Edited by James Hanken and Brian K. Hall. *University of Chicago Press*: 1993. **Vol.1: Development.** Pp. 587. £29.95, \$37.50 (pbk); £59.95, \$75 (hbk). **Vol.2: Patterns of Structural and Systematic Diversity.** £27.95, \$34.50 (pbk); £59.95, \$75 (hbk). **Vol.3: Functional and Evolutionary Mechanisms.** £21.95, \$27.50 (pbk); £55.95, \$70 (hbk).

SKULLS are unfashionable, but they are rewarding subjects for study because their complex structure holds so much easily accessible information, and because they are so closely associated with the brain, major sense organs and feeding. They are splendid material for taxonomic or evolutionary studies, from both phylogenetic and functional perspectives.

Because of these qualities, skulls used to figure large in university zoology courses. Thirty-five years ago, when I was a young lecturer, my students spent a lot of their final year in the museum of the zoology department, learning to identify skulls. That, at least, is what the less enterprising students did. The more enterprising ones spent their time poking long pencils through the skulls' occiputs, writ-

ing the name of each species on the inner surface of the skull roof. We lecturers, of course, spent the evening before the practical examination with long pencils of the kind that have a small eraser on the end, rubbing the names out.

While skulls have been unfashionable, there have been few books about them. The editors of this new publication look back to G. R. de Beer's *The Development of the Vertebrate Skull*, published in 1937, as the great landmark in their subject. There have been important books on skulls since then but no comprehensive treatise. Hanken and Hall regard their work as a complement to de Beer's, not a successor. Interests have changed, so it concentrates on different topics.

This is an advanced book that will be valuable to research workers but will be referred to only occasionally by undergraduates. The three volumes contain 29 long chapters, written by an impressive team of highly appropriate authors.

The first volume covers skull development, and starts with basic embryological questions. In the first chapter R. P. Elinson and K. R. Kao tackle the question of how a radially symmetrical egg becomes a bilaterally symmetrical embryo (an event that must occur before a skull can form). Next, A. G. Jacobson takes up the subject of head segmentation, a topic that I (in my ignorance) thought had been dead since the days of E. S. Goodrich. Jacobson

shows how seven pairs of somitomeres can be seen by the educated eye, anterior to the first somite in embryos of amniotes and teleosts. These chapters are for me the most interesting in the first volume, but there are gems in others. For example, in their chapter on metamorphosis, C. S. Rose and J. O. Reiss tell how larval male ceratioid angler fish develop a new set of jaws to attach themselves as parasites to females, while the females develop their original jaws into a formidable fish trap. The volume ends with a bibliography compiled by the editors, listing 2,674 publications on skull development published between 1937 and 1989. The editors point to electronic databases as sources for more recent references.

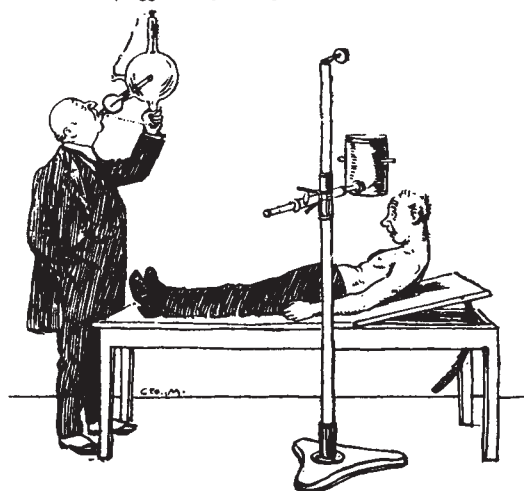
The second volume is about the diversity of skull structure. It begins with a chapter by C. Gans on the evolutionary origin of the skull, based largely on his own hypothesis that hydroxyapatite was first laid down as a dielectric enhancing the sensitivity of electroreceptors: unfortunately, he offers little supporting evidence here. This is followed by chapters by K. S. Thomson on problems of homology and by W. A. Beresford on the skeletal materials of the skull, the latter limited to histology with no discussion of mechanical properties.

Following these introductory chapters, the main part of the volume is devoted to reviews, varying in length and coverage, of the skulls of major vertebrate taxa. R. L. Zusi's chapter on bird skulls is outstandingly good. He discusses the mobility of bird skulls (especially of the mechanism that raises the upper jaw) when the mouth opens; the adaptive radiation of beaks and skulls; convergence, for example between starlings and some shorebirds, both of which probe the ground with their bills and have evolved the capacity to open the bill forcibly against the resistance of the soil; and sexual ornaments such as the casque of the cassowary. L. Trueb reviews the skulls of modern amphibians, illustrating sixteen genera from each of the three orders. By contrast, H.-P. Schultze's chapter on jawed fishes has very little on the teleosts, ignoring the fascinating adaptive radiation of their skulls. And M. J. Novacek describes his chapter on mammal skulls as merely a footnote to de Beer's treatment of the same topic.

The final volume, on functional and evolutionary mechanisms, is closest to my own interests. The authors are excellent, but they do not seem to have been given a clear brief. S. L. Sanderson and R. Wassersug supply a chapter on suspension feeding that says remarkably little (given the context) about skulls. R. E. Lombard and T. E. Hetherington's chapter on hearing and sound transmission is similarly connected only loosely with the main topic of the book. Marvalee Wake bravely attempts a chapter on the skull as a

Phoque Rouge Model.

(Suggested by Prof. Jean Lemarcheur).



The chief feature of this superb machine is the ease by which it may be removed from the stand, and used in any position (as illustrated). The only precaution to be observed is that the operator should be always at a slightly lower level than the tube.

Price, with supply of Water	£14 17 3
Do. do. Fine Old Hungarian Sherry	£17 13 4
Do. do. Oxo	£3 14 7
Special. The patent Whisky & Water-cooled pattern	£314 7 0
Do. with Patient-proof Safe	£3147 14 3

The Department of Half-Seas Over Trade writes: "It is an extraordinary tube."

"Gas bulb surrounded by profanity" — an early nineteenth-century advertisement issued by Cuthbert Andrews, for many years a well known supplier of X-ray apparatus to the medical profession. Taken from Richard F. Mould's *A Century of X-Rays and Radioactivity in Medicine*, which contains more than 700 illustrations, many of them early photographs. *Institute of Physics*, £29.95, \$59.50.

locomotor organ, and then has to cheat a good deal to fill the pages. There are problems of demarcation, with discussions of skull mobility in birds and reptiles occurring both here and in volume two, and the protrusile mechanisms of fish jaws in neither.

It is a remarkable achievement on the part of the editors to have brought out all three volumes simultaneously, so perhaps we should not complain too much about uneven coverage and poor coordination between chapters. Many of us will be glad to possess this work, but it will not always tell us what we hope to learn from it. □

R. McNeill Alexander is in the Department of Pure and Applied Biology, University of Leeds, Leeds LS2 9JT, UK.

Criminal portraits

Chris Brand

The Psychology of Crime. By Philip Feldman. Cambridge University Press: 1993. Pp. 526. £37.50, \$64.95 (hbk); £15.95, \$19.95 (pbk).

OSTENSIBLY, criminality presents an easy target for psychology. Crimes are relatively objective social facts — not conjured up by experts, but reported mainly by members of the public and, when serious, tried by socially representative juries. Crime has a highly skewed distribution: more than fifty per cent of crimes are committed by less than five per cent of the population. Ninety-five per cent of crimes involve males aged between 14 and 40. Crime rates show big group differences, with Japan and Switzerland the safest countries; and crime has increased enormously since 1950 in parallel with rising affluence and the number of criminologists. Yet, as Feldman's textbook faithfully records, from American and British work, there is little agreement about the psychology of crime.

To be sure, the once radical ideas of criminals as victims of loveless homes, poverty and 'labelling' can now largely be set aside. Most of the siblings of criminals have no convictions; self-reported delinquency bears no relationship to social class in modern studies; and police 'targeting' of Black youths barely exceeds the reasonable requirements of economy in the use of manpower. Feldman proceeds fairly and factually to reprove once-fashionable Marxist notions of crime as a form of class war: for crime is as much a scourge of municipal housing estates as it is of leafy garden suburbs; and criminals, victims and the police show excellent agreement about the gravity of crimes and their proper punishment.

Nor has positive explanatory progress

been lacking. While social-environmental accounts of crime remain popular only with tenured sociologists, the psychogenetic methods of twin studies have repeatedly shown a substantial involvement of genetic factors, as well as a minor but clearly detectable involvement of genetic-environment interaction. The problem, however, is to say what exactly is inherited, and which variables 'interact' with each other.

Feldman summarizes some of the modern evidence linking criminality to enduring characteristics of early origin such as aggressiveness, impulsivity, sensation-seeking, hostility, neuroticism and lack of empathy. In particular, he largely admits the evidence for the causal involvement of low IQ in criminality, enabling the startling Black-White difference in crime rates to be handled unprovocatively. Yet trait names are hardly very interesting without an accompanying theory — for example, of the kind there used to be about 'extraversion' (now linked only to delinquency and not to adult criminality) and its basis in poor attention and arousal. Although Feldman warms to the 'interactive' idea that some children are particularly hard to socialize and lack the crucial ingredient of supervision in homes where the father is an alcoholic bully and the mother a pill-swallowing hysteric, he never really commits himself to particular stories that spell matters out. His own theoretical preference is for a 'cognitive-behavioural' approach. But this involves mainly jargon-infected common sense ("stimuli associated with successful property offenses, e.g. opening car doors, will become cues for future theft behavior"); and so little modern empirical work is reported that the chapter on the approach is embarrassingly insubstantial and dated.

Feldman can hardly be blamed for the modern psychology of personality being richer in dimensions than in explanations. Yet it is strange that he will not even ponder such cases as, say, the rape, in 1992, of a hundred-year-old woman in her own bathroom by a sixteen-year-old from Great Yarmouth, England (finally sentenced to nine years in a young offenders' institution). Such cases, with their horrific mixture of sex and violence, should remind the psychologist of the elemental forces and incandescent hostility that are the essential ingredients of many criminal acts. It is not actually easy to burgle or vandalize, let alone to rape and maim, while embarked only on the well-laid schemes of self-reward specified by Feldman's favourite 'rational choice theory'. Although the authoritarian (and 'racist') views and lifelong irreligiosity of criminals are legendary, Feldman is ideologically reluctant to use emotional blunting and unintelligent tough-mindedness as the psychologist's guide in accounting for crime.

Thereby hangs another tale. For, as well as failing to highlight affectionlessness, Feldman has little to say about the crucial interaction between affectionlessness and consciencelessness (or, in some cases, alcohol consumption) that must complete any psychologist's first sketch of the crime scene. All this is strange: for most social scientists have long subscribed to the view (first put forward by Sir Cyril Burt in 1925) that crime and delinquency involve a number of variables in interaction with each other. What seems to have happened is that 'interactionism' has become a form of piety, notably sparing the faithful from the heresy of hereditarianism. Instead of looking for actual multiplier effects between the psychological and social variables known to be implicated in criminality, psychologists have settled for seeing off sociology. Too faint-hearted to invoke the principal potentiating forces that Freud had called — perhaps too romantically for psychologists with scientific aspirations — *eros* and *thanatos*, psychologists thus limit themselves to latter-day behaviourism. Instead of a psychology of crime, Feldman fills out his book with chapters on policing, the courts, the penal system and crime prevention. This will be useful enough to some students of crime, as will the detail on the complete failure, so far, of psychological 'therapies' for crime; but readers wanting some 'psychology' to apply to Feldman's facts and figures will have to supply it themselves. □

Chris Brand is in the Department of Psychology, University of Edinburgh, George Square, Edinburgh EH8 9JZ, UK.

Bird watching

Sara J. Shettleworth

Vision, Brain and Behaviour in Birds. Edited by Philip Zeigler and Hans-Joachim Bischof. MIT Press: 1993. Pp. 415. \$75, £67.50.

BIRDS are among the most visually accomplished of animals. The peacock's tail and the bowerbird's elaborately decorated bower testify to the importance of visual signals in birds' lives. Sexual selection may have even worked on some birds' ability to respond to such subtle visual features as symmetry of body pattern.

Birds are also masters of coordinating vision and movement. Consider a swallow capturing insects while flying, a heron standing on the shore and spearing underwater prey, or a gannet plunging out of the sky for fish. Even birds that feed on stationary prey, such as the granivorous pigeon, can exquisitely adjust the width of their gape to the diameter of the seed they

are aiming to peck. Birds' larger-scale feats include long-distance migration and homing, still mysteries in many ways but controlled at least partly by vision, and long-lasting memories for locations of stored food by titmice, nutcrackers and a few other species. Finally, the fact that complex visual behaviour has evolved in a group widely separated from the better-studied mammals raises important comparative questions about the way in which anatomically and phylogenetically distinct

develops as a consequence of the embryo's position in the egg.

This book covers almost every aspect of vision and visually guided behaviour in birds, from the structure of the eye, colour vision and the functional and developmental anatomy of the visual system, to prey capture by herons and kingfishers, laboratory tests of visual-concept discrimination and the neurobiology of homing and spatial memory. All 22 chapters are written by experts in their fields. Although the coverage is broad, the volume does not pretend to be encyclopaedic. Rather, each chapter is a clear and well-organized introduction to a particular topic, while the extensive and largely up-to-date reference lists can lead the

reader into greater depth. Anyone studying vision in species other than birds is likely to find material of comparative interest here, and the book should be read from cover to cover by students of any aspect of bird behaviour or brains. Those studying birds' visually guided behaviour will become more aware of their subjects' visual capabilities, such as some species' possession of ultraviolet sensitivity and two foveas; and those investigating basic visual processes in birds will learn more about what their subjects may be using vision for. □

Sara J. Shettleworth is in the Departments of Psychology and Zoology, University of Toronto, Ontario, Canada M5S 1A1.

Cyril Webster/NHPA

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

A male satin bowerbird (*P. violaceus*)

species have solved their common problems of interacting with the visual world.

Most laboratory work on the adult avian visual system and visually guided behaviour has used the readily available and robust pigeon. Indeed, 'pigeon' is omitted from the index of species names in the book on the grounds that almost every chapter mentions this animal. From work on pigeons, and the almost equally popular domestic chicken, one can learn the answers to questions applicable to many avian species. Why do some birds bob their heads when they walk, rather than moving them smoothly forward in line with the body? How does an animal with eyes on the sides of its head integrate information from separate visual fields?

Attempts to answer such questions have produced some surprises. For example, not only does information from the lateral visual fields not readily transfer from one hemisphere to the other, but in some cases there may also be hemispheric asymmetry, with one 'eye system' specialized for spatial and the other for object-specific information. In chicks, this lateralization

Vulcans, terminators and science

Harry Kloor and Dennis Harp

Fantastic Voyages: Learning Science Through Science Fiction Films. By Leroy W. Dubeck, Suzanne E. Moshier and Judith E. Boss. *American Institute of Physics/Oxford University Press: 1994.* Pp. 327. \$40, £25.

SCIENCE fiction has an enormous influence on young people's perceptions of science and scientists. So it seems sensible to use it as an aid in teaching science. The thread tying science to science fiction forms the fabric of *Fantastic Voyages*, a science literacy text for nonscientists.

Only abbreviated descriptions of the principles of physics and biology can be presented in such a book. Although most of the coverage of science is easy to read, it is rather dry and will probably entertain only the most dedicated reader. But the physics section does include helpful and sometimes humorous accounts, and nonscientists will be thrilled to learn that it is 99.9 per cent formula-free; readers will obtain a basic grasp of Newton's laws of motion, special relativity, diffusion and thermodynamics. Certain chapters are scientifically weak though, most notably those on electricity and magnetism and evolution, and the accuracy of complex concepts is often compromised by oversimplification.

The essence of the book lies in the illumination of science through selected films, but instead of using the rich archive to provoke an interest in the wonders of science, the authors have allowed stodgy and overly critical analysis to prevail. Although undeniably a matter of taste, admirers of science fiction will on the whole be disappointed, and many would-be followers of science will be turned off by the 'righter than thou' approach.

In many of the examples the physics author seems to have missed the concep-

tual boat. For instance, when Yoda uses his mind to lift a spacecraft in *The Empire Strikes Back*, he is trying to control an etheric, perhaps omnipotent substance that pervades all creation, rather than generate all the energy involved himself. And because Superman is often shown carrying an immensely heavy object while flying, why can't he support one without it pushing him through the rocks on which he is standing?

Nevertheless, the idea of using science fiction films to convey science in an interesting way is sound, and anyone preparing a course using the genre should at least consult this book. □

Harry T. Kloor and Dennis Harp are in the National Science Outreach Program, Purdue University, West Lafayette, Illinois 47907, USA.

■ Aficionados and newcomers will find both literary and scientific inspiration in *The Norton Book of Science Fiction* edited by Ursula K. Le Guin and Brian Attebery (Norton, \$27.50, £22). This 869-page anthology contains 67 stories by North American writers published between 1960 and 1990. In her introduction, Le Guin argues that science-fiction cannot be judged purely on the merits of its scientific content: "there are serious and beautiful science-fiction stories in which the science is completely imaginary and the technology not only implausible but impossible. . . The great boondoggle, popularly and accurately known as 'Star Wars', is an instance of governmental sci fi — a rationally conceived but technologically absurd program based on wish-fulfillment".

Correction

A typographical error crept into the review of *Chromosomes: A Synthesis* in last week's issue (*Nature* 368, 26; 1994). The reviewer's name is Joy Delhanty. Our apologies for the mistake.

Single myosin molecule mechanics: piconewton forces and nanometre steps

Jeffrey T. Finer^{*}, Robert M. Simmons[†] & James A. Spudis^{†*}

^{*} Departments of Biochemistry and Developmental Biology, Beckman Center, Stanford University School of Medicine, Stanford, California 94305, USA

[†] MRC Muscle and Cell Motility Unit, Randall Institute, King's College London, 26–29 Drury Lane, London WC2B 5RL, UK

A new *in vitro* assay using a feedback enhanced laser trap system allows direct measurement of force and displacement that results from the interaction of a single myosin molecule with a single suspended actin filament. Discrete stepwise movements averaging 11 nm were seen under conditions of low load, and single force transients averaging 3–4 pN were measured under isometric conditions. The magnitudes of the single forces and displacements are consistent with predictions of the conventional swinging-crossbridge model of muscle contraction.

DURING muscle contraction, chemical energy from ATP hydrolysis is converted to the relative sliding of actin and myosin filaments and force production. Although the actomyosin system has been extensively studied, the mechanism underlying its mechanochemical energy transduction remains unknown. The conventional swinging-crossbridge theory^{1,2} proposes that for each ATP hydrolysis, myosin binds to actin and undergoes a conformational change or power stroke before subsequently detaching. This theory assumes that the myosin step size, or movement of actin relative to myosin for each ATP hydrolysis, is less than 40 nm, a value limited by the physical dimensions of the myosin head. The solution of the crystal structure of the myosin head led to the suggestion that a long α -helical domain might serve as a lever-arm capable of producing a movement of about 6 nm by a conformational change³. Early mechanical studies on muscle fibres estimated the myosin step size to be well within the structural constraints of the conventional theory^{4,5}. There have been conflicting estimates of the myosin step size from *in vitro* motility assays, however: some measurements^{6–8} are consistent with the conventional theory, whereas others^{9,10} suggest much larger step sizes which are incompatible with the theory. There is also evidence from mechanical studies on muscle fibres that challenges the conventional theory by suggesting that myosin may be capable of undergoing multiple power strokes for each ATP hydrolysis^{11–15}.

These inconsistencies arise because each estimate of the step size is based upon a number of uncertain assumptions, particularly concerning the number of myosin heads involved in a given measurement. The duty ratio, or the fraction of the myosin ATPase cycle that is spent in a force-generating or strongly bound state, has been estimated to be as low as 5 per cent⁷ or as high as 60–90 per cent^{10,13} under sliding conditions in *in vitro* motility assays. As a result, the estimated number of myosin molecules involved in a measurement could easily differ by an order of magnitude. Varying estimates have also been obtained of the number of myosin heads bound to actin or producing force in isometrically contracting muscle fibres as compared with rigor. Under physiological conditions, stiffness measurements suggest that 75 per cent of the myosin molecules are bound to actin¹⁶; X-ray diffraction intensity measurements give values as high as 90 per cent¹⁷; and electron spin resonance measurements suggest that 24 per cent of the myosin heads are bound to actin in a rotationally disordered state¹⁸ and 12 per cent in a rigor-like state¹⁹. It has therefore become clear that in order to resolve these issues, it is necessary to develop new techniques with the resolution necessary to probe the mechanical properties of myosin at the level of single molecular events.

Towards this goal, actomyosin *in vitro* motility assays have become increasingly sophisticated. Unlike kinesin–microtubule assays, in which single kinesin molecules can support continuous movement of microtubules²⁰, actomyosin motility assays are complicated by the comparatively large fraction of time in the ATPase cycle that myosin spends detached from actin. During this time, the actin filament will diffuse away from a myosin molecule unless it attaches to another molecule in the vicinity. This difficulty has been overcome by including in the assay buffer the inert polymer methylcellulose to prevent actin from diffusing away⁷, which made it possible to measure the velocity of actin filaments moving on only a few myosin molecules⁸. It is also possible now to measure the force on single actin filaments moving on a myosin-coated surface either by measuring the bending of a glass microneedle attached to one end of an actin filament^{13,21} or by attaching a latex bead to the actin filament and measuring its movement in a single-beam gradient optical trap²². The optical trap technique has been used to measure single molecule events by kinesin molecules^{23–25} and it offers advantages over microneedles, such as ease of use and the ability to change the stiffness of the trap in the middle of an experiment.

Here we extend the use of the optical trap technique in motility assays to measure nanometre movements and piconewton forces at millisecond rates²⁶. By adding a second optical trap an actin filament can be held and manipulated through beads attached to each end of the filament (Fig. 1a), which prevents the filament from diffusing away from surfaces sparsely coated with myosin. By placing myosin molecules on a support above the coverglass surface, interactions of either the actin filament or the latex beads with the surface are minimized. By using an electronic feedback system, it is possible to measure forces under approximately isometric conditions.

Single displacements

The basic design of the experiment involved the firm attachment of silica beads to a microscope coverslip to provide docking platforms for myosin molecules (Fig. 1a). This coverslip was used to construct a flow cell as in the myosin-coated surface *in vitro* motility assay²⁷. The coverslip was then coated with skeletal muscle heavy meromyosin (HMM) at a density that was insufficient to support continuous movement of actin filaments. Polystyrene beads coated with *N*-ethylmaleimide (NEM)-treated HMM were attached to actin filaments and applied to the flow cell in the presence of ATP. An actin filament with a bead attached near each end was caught and held in mid-solution with two optical traps. The brightfield image of one of the beads was projected onto a quadrant photodiode detector for high resolution position detection. For displacements of the bead up to ~200 nm away from the centre of the trap, the bead position

† To whom correspondence should be addressed.

is proportional to an applied force acting on the bead, so that it is possible to measure force as well as displacement²⁶. The actin filament was pulled taut by moving the second bead until the force on the actin filament was ~ 2 pN, and it was then brought close to the surface of the coverslip so that it could interact with one or a few HMM molecules on the silica bead support. The feedback enhanced laser trap is shown schematically in Fig. 1b.

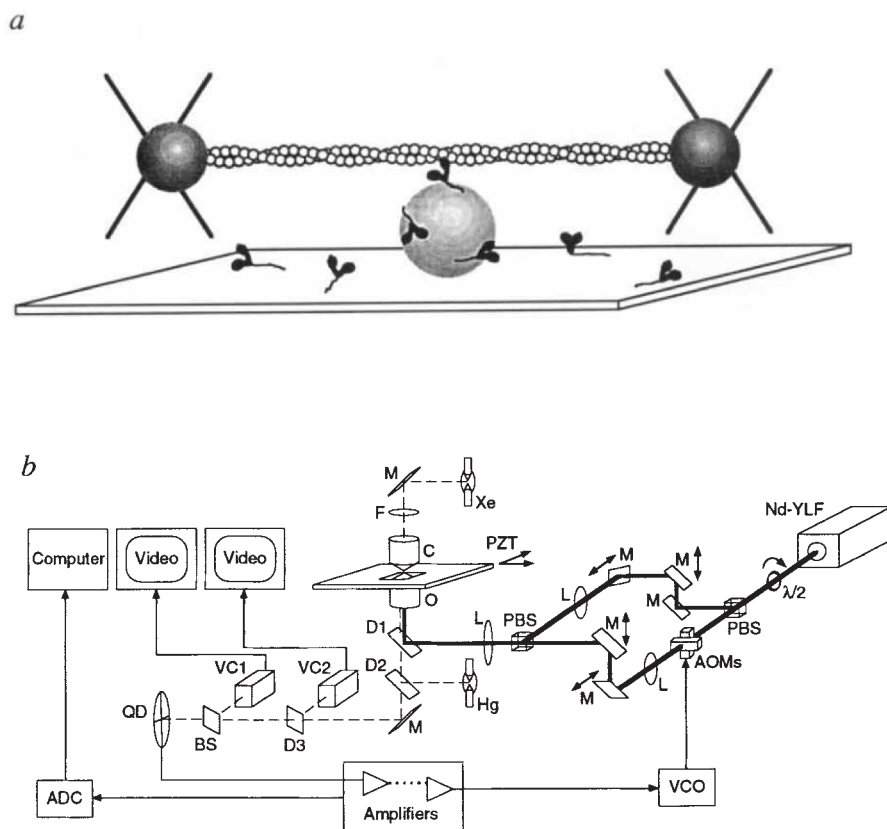
To study myosin movement under conditions of low load, it was necessary to use relatively compliant optical traps. The stiffness of each trap (0.02 pN per nm; total stiffness opposing motion, 0.04 pN per nm) was chosen to be as large as possible to decrease the amplitude of the brownian motion of the beads while not exceeding a force that would impede a myosin

molecule from producing its full displacement. When the actin filament was brought in contact with a sparsely coated surface on the silica bead support, the trapped bead attached to the actin filament showed rapid transient movements in the direction along the actin filament (Fig. 2a, top trace), but not in the direction perpendicular to the actin filament (Fig. 2a, bottom trace). These displacements were almost without exception in one direction, presumably corresponding to the polarity of the actin filament. The leading and falling edges of the displacements were typically as fast as the response time of the measuring system (a few milliseconds).

At saturating ATP concentration (2 mM), the average size and duration of the single displacements were 12 nm and ≤ 7 ms, respectively (Table 1). The duration was close to the resolution

FIG. 1 a, Schematic diagram (not drawn to scale) illustrating the use of two optical traps that are focused on beads attached to a single actin filament, which is held near a single HMM molecule. The filament is pulled taut and lowered onto a silica bead that is firmly fixed to a microscope coverslip and sparsely coated with HMM. b, Schematic diagram of the feedback enhanced laser trap. The thick solid lines represent the optical trap paths; the dashed lines represent imaging paths. Optics include a half-wave plate ($\lambda/2$), polarizing beam splitters (PBS), mirrors (M), lenses (L), dichroic filters (D1-3), microscope objective (O), condenser (C), filter (F), and beam-splitting mirror (BS). Two orthogonal acousto-optic modulators (AOMs) driven by voltage-controlled oscillators (VCO) are used for rapidly deflecting the laser beam in two dimensions. The bead image is projected onto a quadrant photodiode detector (QD) and displacements of the bead in two dimensions, Δx_B and Δy_B , are obtained from the differential outputs of the photodiode elements. The dotted line represents a feedback loop which is closed in order to make measurements under isometric conditions, in which case the quadrant detector outputs are fed through amplifiers to the drivers for the AOMs. A substage consisting of piezo-electric-transducers (PZT) is used for calibrating force (Fig. 4). This system is described in more detail in ref. 26.

METHODS. Silica beads (1 μ m diameter; Bangs Laboratories) were firmly fixed to a microscope coverslip by suspending them in 0.05% Triton X-100 and spreading them onto the coverslip, which was then air-dried. The surface was then either coated with nitrocellulose⁴¹ or siliconized by treatment with 0.2% dichlorodimethylsilane (Dow Corning, Z1219) in chloroform. The coverslip was then used to construct a flow cell as described²⁷. Rabbit skeletal muscle HMM (1–5 μ g ml⁻¹) was applied to the flow cell for 2 min. Polystyrene beads (1.0 μ m MX Covaspheres, Duke Scientific), labelled with NEM-HMM as described⁴², were mixed with rhodamine-phalloidin-labelled actin filaments at a ratio of ~ 1 bead to 10 μ m F-actin before being applied to the flow cell in motility buffer (25 mM KCl, 4 mM MgCl₂, 1 mM EGTA, 10 mM DTT, 25 mM imidazole, pH 7.4) at 21 °C. Two beads were placed on an actin filament near its ends (typically 5–10 μ m apart), with each bead held in an optical trap. The most successful technique was to use one optical trap to hold a bead with a filament already attached. The actin filament was then straightened by a solution flow produced by the motorized stage controls, and a second bead manoeuvred by the second optical trap was attached to the free end of the filament. The filament was then pulled taut by moving the second trap with a motorized mirror under computer control, and the filament was lowered onto the silica bead using the fine focus of the microscope objective. The optical trap design used a Nd:YLF diode-pumped laser (Spectra-Physics, TFR; 1.047 μ m) and a custom-built inverted microscope with a high numerical aperture objective (Zeiss, 63 $\times$ Planapochromat, 1.4 NA, DIC grade).



Two traps were produced by splitting the laser beam before the AOMs using a half-wave plate ($\lambda/2$) followed by a polarizing beam splitter (PBS). The traps were the same strength (12 mW, measured before the objective). One trap position was controlled for small, fast movements (resolution < 1 nm; response time about 10 μ s) using the AOMs (Isomet, 1206C), and both trap positions were controlled for larger, slower movements using d.c. motors (Newport, 860A-1-HS) to move mirrors. Specimens were observed simultaneously by brightfield and epifluorescence using video cameras (VC1-2) using a 75 W xenon arc lamp (Xe) and a 100 W mercury arc lamp (Hg), respectively. Dichroic filters (D1-3) were used to separate the laser light, the brightfield illumination, and the epifluorescence. (Transmission: D1, 450–1,000 nm; D2, > 565 nm; D3, 700–1,000 nm; F, > 700 nm; BS, 95% 700–1,000 nm.) The position of the bead in the trap equipped with AOMs was monitored by the quadrant photodiode detector (QD) (Hamamatsu, S1557). Δx_B and Δy_B were obtained by appropriate additions and subtractions of the four quadrant outputs. The bandwidth was 100 Hz, limited by the 200 M Ω resistors used in the current-to-voltage converters and stray capacitance. Data was sent through an analog-digital converter (ADC) and stored on a PC computer (see Fig. 2). The quadrant detector was mounted on micrometers, and these were used to determine the sensitivity of the detector and the AOMs.

TABLE 1 Average single myosin molecule measurements

ATP concentrations (mM)	Single displacements (nm)	Displacement durations (ms)	Single forces (pN)	Force durations (ms)
2	12 ± 2.0 (22)	≤ 7 (22)	3.4 ± 1.2 (85)	18 ± 6 (85)
0.01	11 ± 2.6 (36)	72 ± 29 (36)	3.5 ± 1.3 (43)	25 ± 9 (43)
0.001	11 ± 2.5 (21)	260 ± 140 (21)	3.4 ± 1.4 (50)	190 ± 150 (50)

Values are mean ± standard deviation (*n*). The differences between force and displacement durations were examined for statistical significance, using error estimates obtained from the distribution of the means of experimental runs. At 2 mM and 10 μ M ATP these differences were highly significant ($P < 0.01$), but at 1 μ M ATP the difference was not significant ($P > 0.05$).

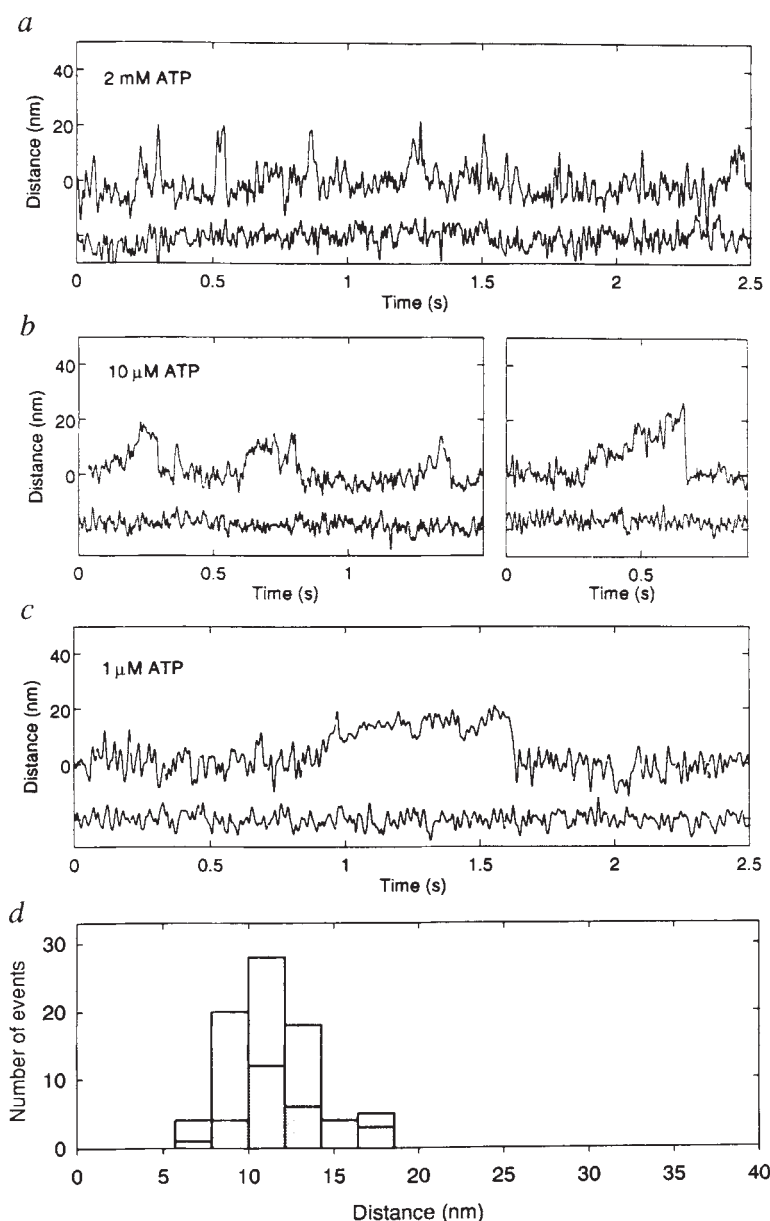
of measurement, and we therefore lowered the ATP concentration to 1 μ M or 10 μ M to delay the dissociation of myosin from actin. These concentrations are well below the apparent K_m ($\sim 50 \mu$ M) for the sliding velocity²⁷. The HMM density was also decreased, to the point where many of the actin filaments tested showed no transient displacements, so that when interactions were detected they most likely involved only one or a very few HMM molecules. Single displacements were detected above the noise at 10 μ M ATP and 1 μ M ATP (Fig. 2*b, c*). The size of the steps, 11 ± 2.4 nm (mean $\pm$ s.d.), was the same at both high and

low ATP concentrations, but the average duration of the displacements increased as the ATP concentration decreased (Table 1). The peak amplitude distribution was independent of total trap stiffness over the range 0.014–0.08 pN per nm, implying that the step size is nearly independent of load in this range. This experiment was done using siliconized surfaces and nitrocellulose-coated surfaces, with similar results (Fig. 2*d*). It was noticeable in many records that the brownian noise was markedly reduced during the steps, presumably because of the increased stiffness associated with the HMM–actin link²⁵. In some records it appeared that the rise time of the displacements (Fig. 2*b, c*) was relatively slow, but examination of all of the records failed to show a consistency in this pattern. This point requires further investigation.

When the HMM density on the surface was increased to levels just below that which supported continuous movement of actin over larger distances, the trapped bead no longer returned to its baseline position after each displacement. Instead the bead moved for distances larger than a single step before abruptly returning to the baseline (Fig. 3*a*). At the trap stiffness used in these measurements, a free bead returns to its equilibrium position with a time constant of < 1 ms²⁶, so that each rapid return to the baseline in these experiments probably corresponded to

FIG. 2 Single displacements at low load with *a*, 2 mM; *b*, 10 μ M; and *c*, 1 μ M ATP. The upper trace in each record shows the movement of the trapped bead in the direction along the actin filament. The lower trace in each shows the movement of the bead in the perpendicular direction. The HMM density was highest in *a* and lowest in *c*. *d*, Distribution of single displacement amplitudes obtained from data at 2 mM, 10 μ M and 1 μ M ATP. The distribution shows data obtained from both siliconized surfaces (white) and nitrocellulose surfaces (shaded). Each trap: laser power 12 mW, stiffness 0.02 pN/nm.

METHODS. The HMM concentrations applied to the flow cell were: 5 μ g ml⁻¹ (*a*), 2 μ g ml⁻¹ (*b*) and 1 μ g ml⁻¹ (*c*). ATP concentrations of 1 μ M or 10 μ M were maintained by the presence of 1 mM phosphocreatine and 0.1 mg ml⁻¹ creatine phosphokinase. The laser trap stiffness was calibrated by applying viscous forces to a trapped bead as described in Fig. 4. Signals were sampled at 4 kHz (R. C. Electronics, ISC-16), recorded on a computer, and subsequently filtered during analysis using up to 8-point averaging, and a 2–8-point Hanning filter. Essentially the same results were achieved using steep-cut Bessel filters. Single displacement amplitudes were measured, using the minimum filtering possible, by fitting lines through the baseline noise on either side of an event, and (where possible) through the noise at the plateau of the event. Not all of the displacements shown here were scored for inclusion in the distribution shown in *d* because many did not meet the following criteria: the displacements had to be isolated events; the baseline brownian motion had to be at the same level on either side of the peak; and the displacements could not appear similar in character to the multiple displacements shown in Fig. 3. An example of a displacement that was not scored for the latter reason is shown in *b*, right panel. Displacements smaller than or equal to the level of the brownian noise (equivalent to the perpendicular motion shown in the lower trace) were not scored, and as a result the distribution shown in *d* may be truncated below ~ 5 nm. Compliance of the filament was ignored, as it was estimated to be $\ll 0.3$ nm per pN, assuming a Young's modulus of $\gg 6.8 \times 10^8$ N m⁻², a filament length of $\leq 5 \mu$ m, and a cross-sectional area of 24 nm² (refs 43, 44).



a period when there were no HMM molecules bound to the actin filament. This would be expected to occur frequently at the subsaturating HMM densities used in these measurements. In many instances the bead clearly moved in a few discrete steps (Fig. 3*b*, left panel) before returning to the baseline. This feature is quantified in Fig. 3*b* (right panel), which is a histogram that shows the distribution of distances away from the baseline as a function of time spent at each distance. The bead spent long periods of time at discrete levels and moved quickly between levels. The distance between levels averaged 11 ± 3.0 nm (mean $\pm$ s.d.) (Fig. 3*c*), which is consistent with the size of the single movements shown in Fig. 2 and Table 1. As such multiple steps were observed infrequently at very low surface HMM densities, it is likely that they correspond to a small number of HMM molecules attaching to a filament and moving it sequentially.

It would be expected that at high, saturating surface density of HMM molecules, the velocity of filament sliding would be at least as great as the average unitary step divided by its duration. We measured these velocities in a motility assay at the three ATP concentrations, and found them to be $3.0 \mu\text{m s}^{-1}$ (2 mM ATP), $0.26 \mu\text{m s}^{-1}$ (10 μM) and $0.08 \mu\text{m s}^{-1}$ (1 μM). These compare favourably with the calculated velocities from the single displacements of ≥ 2 , 0.15 and $0.04 \mu\text{m s}^{-1}$, respectively. The approximate agreement between the values of velocity observed at high

HMM density and the values calculated from the size and duration of single displacements, strongly suggests that these single displacements are the underlying basis of filament sliding, although precisely what happens during overlapping interactions remains to be investigated.

Single forces

To measure the force produced by HMM molecules, a much stiffer trap was needed than was used to measure displacements. The trap stiffness was increased to 6 pN per nm by means of a feedback system in which the quadrant detector output signals were fed into the driver circuits for acousto-optic modulators which were used for making rapid trap displacements²² (Fig. 1*b*). When an external force in the form of a viscous drag was applied to a trapped bead without feedback, the bead was displaced from the centre of the trap (Fig. 4*a*); this displacement was directly proportional to the force applied (Fig. 4*c*). In contrast, when the feedback loop was closed, the centre of the trap moved by an amount proportional to the applied force and the bead was essentially prevented from moving (Fig. 4*b, c*). Thus, the displacement of the trap position could be used as a measure of force on the trapped bead under approximately isometric conditions.

The procedure for measuring single forces was as follows. When an actin filament was brought in contact with HMM molecules and single displacements were observed, the feedback loop was closed and force fluctuations were then measured. At sufficiently low HMM densities, single force transients were observed (Fig. 5*a-c*). Just as for displacements at low load, forces were nearly always found to be in one direction along a particular actin filament. The magnitude of the forces covered a broad distribution, ranging from 1 to 7 pN, and averaged 3.4 ± 1.2 pN (mean $\pm$ s.d.) (Fig. 5*d*). The distribution of forces was the same at the different ATP concentrations (Table 1). As observed for the single displacements at low load, the duration of the single force transients increased at low ATP concentrations (Table 1). At these low HMM densities, the force transients usually appeared as isolated events, but at slightly higher HMM densities the events occurred very close together and often were additive (data not shown). This is the corresponding process to multiple steps at low load, and presumably is due to the overlapping action of several HMM molecules.

Discussion

The ability to measure the mechanical properties of single myosin molecules provides us with direct evidence that under conditions of low load, myosin undergoes stepwise displacements with an average step of about 11 nm, and under isometric conditions, myosin produces an average force of 3–4 pN. In discussing our results we assume that the single forces and displacements were derived from interactions of actin with single myosin heads, as there is evidence that the two heads on a given myosin molecule act independently^{28,29}. However, we cannot at present discount the possibility that some of the interactions we have observed involve both heads of a myosin molecule. One complication that probably affects the values of the force and displacement in our experiments is that the HMM molecules are randomly oriented on the surface, which is likely to reduce the efficacy of force production or displacement. However, even myosin heads that are in the completely wrong orientation still move actin filaments *in vitro* at one-tenth the normal velocity^{30,31}. It seems likely, therefore, that the true average force and average displacements produced by correctly oriented myosin heads are not higher than the upper end of our range of values. These values are well within the range of those allowed by the swinging-crossbridge model², and the average value of 11 nm agrees with the value of 14 nm found by Ford *et al.*⁵ for the maximum range of rapid recovery after a shortening step. However, the largest displacements that we observed may not fit easily into the specific model suggested by Rayment *et al.*³,

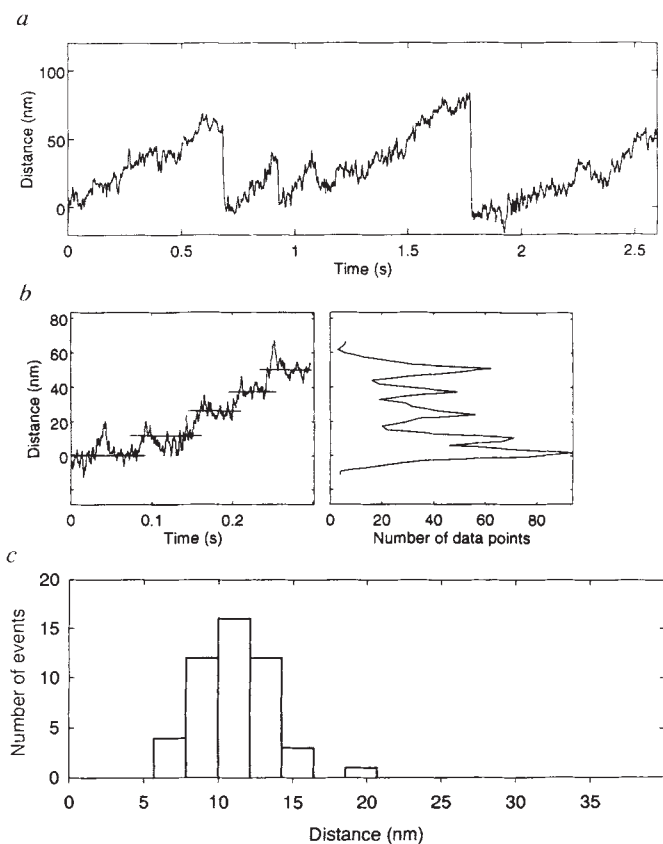
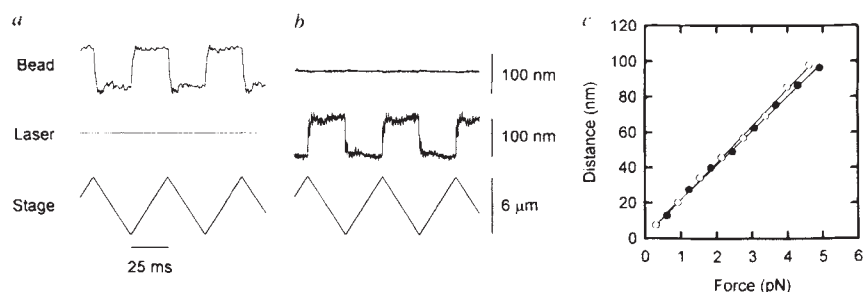


FIG. 3 *a* and *b* (left panel), Multiple displacement records at higher HMM density than for Fig. 2, 10 μM ATP. *b* (right panel), Distribution of distances away from the baseline for the records shown in the left panel. The abscissa is proportional to the time the bead spent at each position along the ordinate. *c*, Distribution of distances between levels for multiple displacement records at 10 μM ATP and 1 μM ATP. Each trap: laser power 12 mW, stiffness 0.02 pN per nm.

METHODS. As described in Fig. 2 legend, but $5 \mu\text{g ml}^{-1}$ HMM was applied to the flow cell. The distribution in *b* (right panel) had a bin size of 2.2 nm. The distribution of steps in *c* was formed by fitting lines through adjacent plateaus and measuring the distance between them.

FIG. 4 Example of optical trap stiffness calibration with and without feedback control. *a*, Without feedback: a viscous force that alternated in direction was applied to a trapped bead by applying a triangular wave (lower trace) to the microscope stage position; the bead position showed an approximately square wave response. *b*, With feedback: as in *a*, but in this case the trap position showed an approximately square wave response, and the bead remained essentially stationary. *c*, Displacement versus force curve. ●, Bead movement without feedback; ○, trap movement with feedback. The trap stiffness under feedback control increased by ~300-fold from 0.05 pN per nm without feedback to 15 pN per nm with feedback.

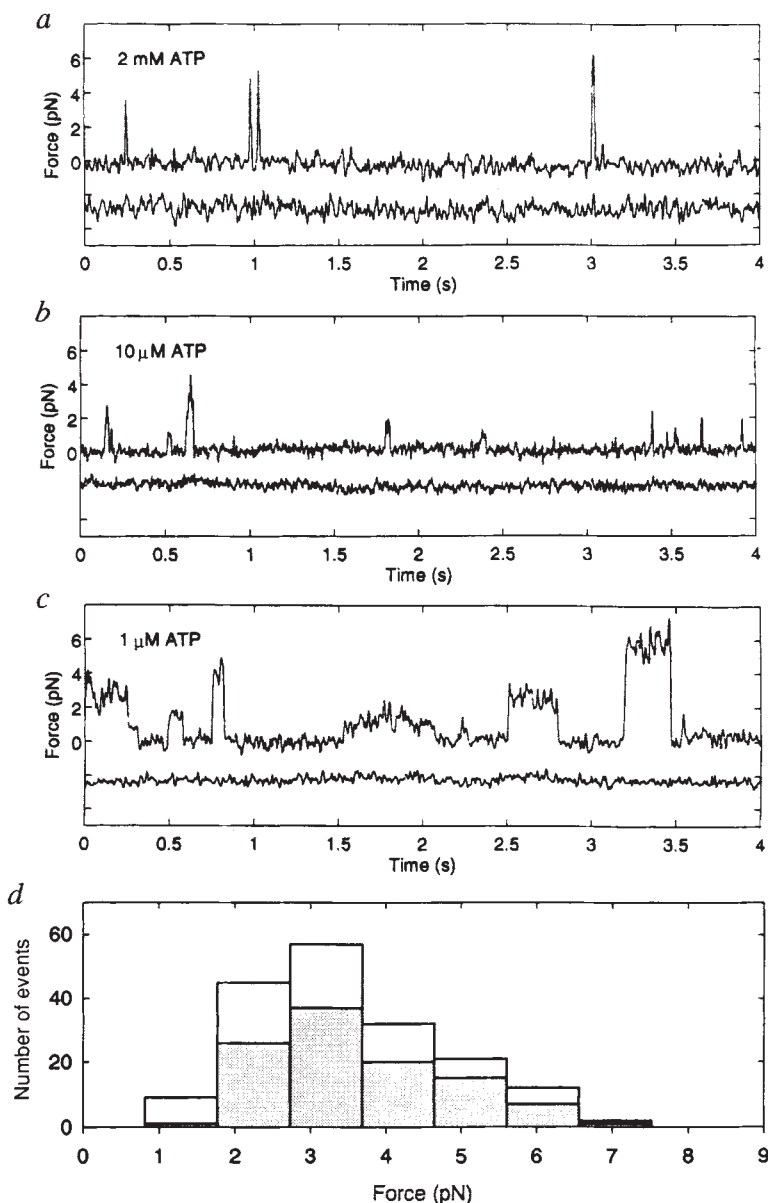
METHODS. A viscous force was applied to a trapped bead 5 μm from the coverslip surface by moving a substage with piezo-electric transducers (Physik Instrumente, P771) to produce a solution flow. Stokes' law was used to calculate the force, which is directly proportional to the velocity of the solution flow²⁶. The small amount of creep in the response was due to hysteresis in the PZT response (not shown). The average value



of displacement was used and shown to be directly proportional to the measured average stage velocity. Under feedback control the bead movement, including brownian motion, was less than 2 nm peak-to-peak. As the displacement was small, the trap stiffness was obtained by averaging 50 traces of the bead response (*b*, top trace). The bandwidth of the feedback system, measured from the response to a small input perturbation, was 830 Hz. The increase in bandwidth compared to that of the detector (100 Hz, single pole) resulted from the high feedback gain.

FIG. 5 Single force transients near isometric conditions with *a*, 2 mM; *b*, 10 μM ; and *c*, 1 μM ATP concentrations. The upper trace in each record shows the forces in the direction along the actin filament. The lower trace in each shows forces in the perpendicular direction. *d*, Distribution of single force amplitudes obtained from data using 2 mM, 10 μM and 1 μM ATP concentrations. The distribution shows data obtained from both siliconized surfaces (white) and nitrocellulose surfaces (shaded). Each trap: laser power 12 mW. Stiffness of trap with feedback, 6 pN per nm; stiffness of trap without feedback, 0.02 pN per nm.

METHODS. Single actin filaments were pulled taut as described in Fig. 1, and feedback control was applied to the bead imaged on the quadrant detector. The HMM concentrations applied to the flow were: 2 $\mu\text{g ml}^{-1}$ (*a*) and 1 $\mu\text{g ml}^{-1}$ (*b*, *c*). The criteria used to select single force transients were similar to those used for single displacements (Fig. 2 legend). Force transients were not scored if they were in the direction that caused the filament to slacken. Calibration for force was made as for Fig. 4, but with the bead centred 1 μm from the surface (equal to the distance of the actin-bound beads from the surface in this experiment). It was found by two methods that the trap stiffness was the same at 5 μm and 1 μm from the surface: First, the same calibration procedure gave 1.4 times greater bead displacement at 1 μm , corresponding well to the raised viscous force near a surface⁴⁵; second, we compared the amplitude of brownian noise at the two depths (as r.m.s. brownian motion is inversely proportional to trap stiffness) and found that the two were identical. Errors in the force calibration were estimated to be less than 10%.



which is based on the S1 crystal structure. The values of the displacements reported here do not support models in which myosin heads are able to generate movement over very large distances^{9–11}. Models in which a myosin molecule makes multiple interactions with an actin filament for one ATP hydrolysis^{12–15} would also be unlikely, because the single displacements and single forces that we observed at very low HMM density did not appear as a rule to be clustered²⁵. Furthermore, our data show that the durations of the transient displacements and forces depend on ATP concentration so that, at least at low ATP, each interaction requires the completion of the ATP hydrolysis cycle. In addition, as discussed in more detail below, the durations are very close to those that would be expected for single ATP hydrolysis cycles. Models proposing that a myosin molecule makes multiple interactions with an actin filament for one ATP hydrolysis would predict that, at low ATP concentrations, multiple interactions would occur before the final step of product release, and they would presumably be similar in duration to the interactions seen at high ATP. Such models would therefore require a marked preponderance of events of very short duration at low ATP, and this was not observed in our preliminary analysis of the data.

Our results for the ATP dependence of single forces and displacements give some preliminary information about the biochemical steps involved and the effects of mechanical strain. The reciprocal of the mean value of duration of single interactions gives the rate constant for the process which terminates the event. For both single forces and displacements, the variation of the rate constant with ATP concentration is consistent with Michaelis–Menten kinetics, with an apparent K_m for ATP of the same order of magnitude as for sliding velocity in a motility assay, $50 \mu\text{M}$ ²⁷. However, we found significant differences between the durations for single forces and those for single displacements (that is, high and low loads) in the case of $10 \mu\text{M}$ and 2 mM ATP, which would be consistent with a lower saturating rate and apparent K_m at high loads. In turn, such a difference would imply that the second-order rate constant for ATP binding is larger at high loads. The values for the second-order rate constant have to be calculated from concentrations of substrate well below the K_m , and at $1 \mu\text{M}$ ATP the values calculated from $1/(Cr)$, where C is the ATP concentration and t is the average duration, are $3.8 \pm 0.4 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ (mean $\pm$ s.e.m.) for displacement and $5.3 \pm 1.6 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ for force. Although the difference between these values is not statistically significant, it is consistent with our argument. These values are within range of the value of $4.0 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ for the acto-S1 ATPase in solution³². A detailed study will be required to confirm these results, but the difference in the values for the rate of ATP binding at low and high load is probably due to a strain-dependent increase in this rate, which is to be expected for a strong-to-weak transition³³. At saturating ATP the situation is reversed, and single forces lasted longer than single displacements, implying that the rate of detachment rises during filament sliding³⁴. The duration of $\leq 7 \text{ ms}$ for single displacements is equivalent to a rate constant of $\geq 140 \text{ s}^{-1}$. This value can be compared directly with the rates that affect detachment in the acto-S1 ATPase, and it probably corresponds to an isomerization between two AM.ADP states preceding ADP release³⁵. The difference between the durations for force and displacement at high ATP can be explained by a strain-dependence of this transition.

If the force exerted by a correctly oriented myosin head is $3\text{--}7 \text{ pN}$, how does this compare with the average force per myosin head produced by a contracting muscle fibre under isometric conditions? It is difficult to make exact comparisons because of differences in the conditions used, but rabbit skinned muscle fibres produce an isometric force of 0.22 N mm^{-2} (ionic strength, 200 mM , $\text{pH } 7.1$, 20°C ; M. G. Bell, J. A. Dantzig and Y. E. Goldman, personal communication). Assuming a filament lattice spacing ($d_{1,0}$) during activation of 38 nm (ionic strength 170 mM , $\text{pH } 7.0$, 5°C)³⁶, the average force is 1.3 pN per myosin

head³⁷. A similar value was found for frog intact muscle near 0°C (force, 0.25 N mm^{-2} (ref. 38); $d_{1,0}$, 35 nm (ref. 39)), and our results would then suggest that only $20\text{--}40\%$ of the heads develop force at any one time, that is, the duty ratio is $0.2\text{--}0.4$. Earlier estimates of the fraction of myosin crossbridges attached at any moment in an isometric contraction were obtained using several techniques by comparison with the rigor state, in which it is thought that all the crossbridges are bound to actin, and with the relaxed state, in which the crossbridges are either not bound or only weakly bound. Measurements on frog intact muscle of fibre stiffness¹⁶ and X-ray diffraction equatorial intensities¹⁷ suggest values for the fraction of myosin heads attached at any moment of 0.75 and 0.9 , respectively, though these estimates may include bound crossbridges not producing force. On the other hand, electron spin resonance measurements on activated, crosslinked myofibrils show that at low ionic strength (45 mM , $\text{pH } 7.0$, 25°C) a fraction of the heads (0.37) are bound to actin in a rotationally disordered state¹⁸, with 0.12 bound in a rigor-like state¹⁹. Measurements of the intensity of the X-ray layer lines also suggest a low fraction of ordered heads, $0.2\text{--}0.3$ (ref. 17). Our results support these estimates.

The duty ratio under low load can be calculated from the durations of the single displacements measured here and the k_{cat} value for the actin-activated HMM ATPase under similar conditions. At low load and 2 mM ATP, the average duration of single displacements was $\leq 7 \text{ ms}$. There is no direct measurement of the ATPase rate under the single molecule conditions of our experiments (this presents a challenge for the future), but under comparable solution and temperature conditions, the k_{cat} for the ATPase would be about 10 per second per head⁴⁰. Thus under these conditions, the upper limit of the duty ratio under low load is 0.07 . This value is consistent with that obtained by Uyeda *et al.*⁷. Therefore under both low- and high-load conditions, the results reported here are consistent with small duty ratios, and are compatible with the predicted large duty ratios^{10,13} only if the ATPase rate in the motility experiments is 100 per second per head, which seems unlikely.

The picture that emerges from this study is of a simple cycle of force production followed by detachment, in which the isometric force per interaction of a correctly oriented myosin head is $3\text{--}7 \text{ pN}$ and the displacement at low load is $8\text{--}17 \text{ nm}$.

Note added in proof: The number of interactions per ATP hydrolysed can be obtained by dividing the useful energy per ATP ($0.5 \times 10^{-19} \text{ J}$)³⁷ by the energy per interaction (estimated as $0.5 \times 3.4 \times 11 \times 10^{-21} \text{ J}$), giving 2.6 . As our values of force and displacement are underestimates, the correct value may be close to unity. $\square$

Received 27 October 1993; accepted 2 February 1994.

1. Reedy, M. K., Holmes, K. C. & Tregear, R. T. *Nature* **207**, 1276–1280 (1965).
2. Huxley, H. E. *Science* **164**, 1356–1366 (1969).
3. Rayment, I. *et al. Science* **261**, 50–58 (1993).
4. Huxley, A. F. & Simmons, R. M. *Nature* **233**, 533–538 (1971).
5. Ford, L. E., Huxley, A. F. & Simmons, R. M. *J. Physiol.* **269**, 441–515 (1977).
6. Toyoshima, Y. Y., Kron, S. J. & Spudich, J. A. *Proc. natn. Acad. Sci. U.S.A.* **87**, 7130–7134 (1990).
7. Uyeda, T. Q. P., Kron, S. J. & Spudich, J. A. *J. molec. Biol.* **214**, 699–710 (1990).
8. Uyeda, T. Q. P., Warrick, H. M., Kron, S. J. & Spudich, J. A. *Nature* **352**, 307–311 (1991).
9. Harada, Y. & Yanagida, T. *Cell Motil. Cytoskel.* **10**, 71–76 (1988).
10. Harada, Y., Sakurada, K., Aoki, T., Thomas, D. D. & Yanagida, T. *J. molec. Biol.* **216**, 49–68 (1990).
11. Yanagida, T., Arata, T. & Oosawa, F. *Nature* **316**, 366–369 (1985).
12. Higuchi, H. & Goldman, Y. E. *Nature* **352**, 352–354 (1991).
13. Ishijima, A., Doi, T., Sakurada, K. & Yanagida, T. *Nature* **352**, 301–306 (1991).
14. Brenner, B. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10490–10494 (1991).
15. Lombardi, V., Piazzesi, G. & Linari, M. *Nature* **355**, 638–641 (1992).
16. Goldman, Y. E. & Simmons, R. M. *J. Physiol.* **269**, 55p–57p (1977).
17. Huxley, H. E. & Kress, M. J. *Muscle Res. Cell Motil.* **6**, 153–161 (1985).
18. Berger, C. L. & Thomas, D. D. *Biochemistry* **32**, 3812–3821 (1993).
19. Fajer, P. G., Fajer, E. A. & Thomas, D. D. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5538–5542 (1990).
20. Howard, J., Hudspeth, A. J. & Vale, R. D. *Nature* **342**, 154–158 (1989).
21. Kishino, A. & Yanagida, T. *Nature* **334**, 74–76 (1988).
22. Simmons, R. M. *et al. in Mechanism of Myofibrillar Sliding in Muscle* (eds Sugi, H. & Pollack, G. H. J.) (Plenum, New York and London, 1993).
23. Block, S. M., Goldstein, L. S. B. & Schnapp, B. J. *Nature* **348**, 348–352 (1990).
24. Kuo, S. C. & Sheetz, M. P. *Science* **260**, 232–234 (1993).
25. Svoboda, K., Schmidt, C. F., Schnapp, B. J. & Block, S. M. *Nature* **365**, 721–727 (1993).

26. Simmons, R. M., Finer, J. T., Chu, S. & Spudich, J. A. *Biophys. J.* (submitted).
27. Kron, S. J. & Spudich, J. A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6272–6276 (1986).
28. Taylor, E. W. *Crit. Rev. Biochem.* **6**, 103–164 (1979).
29. Cooke, R. *Crit. Rev. Biochem.* **21**, 53–118 (1986).
30. Sellers, J. R. & Kachar, B. *Science* **249**, 406–408 (1990).
31. Yamada, A., Ishii, N. & Takahashi, K. *J. Biochem.* **108**, 341–343 (1990).
32. White, H. D. & Taylor, E. W. *Biochemistry* **15**, 5818–5826 (1976).
33. Goldman, Y. E., Hibberd, M. G. & Trentham, D. R. *J. Physiol.* **354**, 577–604 (1984).
34. Huxley, A. F. *Prog. Biophys.* **7**, 225–318 (1957).
35. Taylor, E. W. *J. biol. Chem.* **266**, 294–302 (1991).
36. Brenner, B. & Yu, L. C. *J. Physiol.* **441**, 703–718 (1991).
37. Bagshaw, C. R. *Muscle Contraction*, 98 (Chapman & Hall, London, 1993).
38. Woledge, R. C., Curtin, N. A. & Homsher, E. *Energetic Aspects of Muscle Contraction*, 103 (Academic, London, 1985).
39. Hasegawa, J. C. & Huxley, H. E. *J. molec. Biol.* **77**, 549–568 (1973).
40. Margossian, S. S. & Lowey, S. *Meth. Enzym.* **85**, 55–71 (1982).
41. Toyoshima, Y. Y. et al. *Nature* **328**, 536–539 (1987).
42. Warrick, H. M. et al. in *Methods in Cell Biology* (eds Scholey, J. M.) 1–21 (Academic, San Diego, 1993).
43. Yoshino, S., Umazume, Y., Natori, R., Fujime, S. & Chiba, S. *Biophys. Chem.* **8**, 317–326 (1978).
44. Ford, L. E., Huxley, A. F. & Simmons, R. M. *J. Physiol.* **311**, 219–249 (1981).
45. Svoboda, K. & Block, S. M. A. *Rev. Biophys. biomol. Str.* (in the press).

ACKNOWLEDGEMENTS. We thank S. Chu for help with the design of the prototype instrument that led to this work, and J. Sleep, T. Uyeda, C. Coppin, A. Friedman and Y. Goldman for comments on the manuscript. This work was supported in part by the NIH Medical Scientist Training Program (J.T.F.), the MRC, the Wellcome Trust, the Fulbright Commission, and NATO (R.M.S.), and a grant from the NIH (J.A.S.).

Protein–protein interactions and 5′-splice-site recognition in mammalian mRNA precursors

Jhumku D. Kohtz^{*}, Sharon F. Jamison[†], Cindy L. Will[‡], Ping Zuo^{*§}, Reinhard Lührmann[‡], Mariano A. Garcia-Blanco[†] & James L. Manley^{*}

^{*} Department of Biological Sciences, Columbia University, New York, New York 10027, USA

[†] Department of Molecular Cancer Biology, Department of Microbiology, Department of Medicine, Duke University Medical Center, Durham, North Carolina 27710, USA

[‡] Institut für Molekularbiologie und Tumorforschung, Philipps-Universität, Marburg, Emil-Mannkopff-Strasse 2, D-35037 Marburg, Germany

Exactly how specific splice sites are recognized during the processing of complex precursor messenger RNAs is not clear. Small nuclear ribonucleoprotein particles (snRNPs) are involved, but are not sufficient by themselves to define splice sites. Now a human protein essential for splicing *in vitro*, called alternative splicing factor/splicing factor 2, is shown to cooperate with the U1 snRNP particle in binding pre-mRNA. This cooperation is probably achieved by specific interactions between the arginine/serine-rich domain of the splicing factor and a similar region in a U1 snRNP-specific protein.

SPLICING of a nuclear pre-mRNA involves an ordered assembly of components resulting in the formation of a functional spliceosome. Factors within the spliceosome, including snRNP particles and non-snRNP proteins, mediate the cleavage and ligation reactions required to generate spliced mRNAs. Definition of the splicing pathway and its components has been an area of intense investigation (reviewed in refs 1 and 2). It is still unknown, however, how the factor(s) responsible for initiation of the assembly of spliceosomal complexes recognize(s) specific splice sites in the pre-mRNA. This issue is particularly important in complex metazoan pre-mRNAs, which often contain immense introns and display complicated patterns of alternative splicing (reviewed in ref. 3).

The involvement of snRNPs in splicing was first shown for U1 snRNP, and subsequently for U2, U4/6, and U5 snRNPs (reviewed in ref. 4). Base-pairing between the 5′ end of U1 snRNA and the 5′ splice site is known to be required, but is not sufficient for proper recognition of that site^{5–7}. In addition to U1 snRNP, U5 snRNP is thought to influence 5′ splice site selection by base pairing with sequences in the exon adjacent to the 5′ splice site⁸, and U6 snRNP can also associate with this region^{9,10}. The importance of the U1 snRNP–pre-mRNA interaction early in spliceosome assembly is underlined by the presence of this snRNP in both the commitment complex (in which it is the only known snRNP) and the pre-spliceosomal E and A

complexes^{11–14}. Commitment was originally defined in yeast¹² as a preferential processing of a pre-mRNA substrate incubated in splicing-competent extracts in the presence of excess competitor RNA added later in the reaction. The commitment complex is the first functional pre-spliceosomal complex formed on pre-mRNA that cannot be competed by exogenous pre-mRNA.

Mammalian 5′ splice sites are highly divergent compared with those found in yeast¹⁵, emphasizing the need for factors that stabilize U1 snRNP–pre-mRNA interactions. It is known that recognition of the 5′ splice site by U1 snRNP requires both U1 RNA and U1-specific proteins, including the U1-C protein^{16–18}. Involvement of the protein U1 70K (*M_r* 70,000), another U1 snRNP-specific protein, in the pre-catalytic steps of splicing depends on its state of phosphorylation¹⁹. An association of cryptic 5′-splice-site-containing pre-mRNAs with U1 snRNP can only be detected in splicing extracts, also indicating the need for stabilizing, *trans*-acting factors²⁰. Indeed, multiple non-snRNP RNA-binding proteins have been identified that are essential for splicing and/or can influence splice-site selection. The best characterized include ASF/SF2^{21–24}, SC35^{25,26} and U2AF^{27–29}. U2AF is an essential splicing factor necessary for binding of U2 snRNP to the branch site^{27,28}. The U2AF protein contains RNP-type RNA-binding domains (RBDs; reviewed in refs 30, 31) and an arginine/serine-rich region (RS domain). ASF/SF2 and SC35 also contain RBDs and belong to a larger family of RS-domain-containing proteins, the SR proteins³², that are able to complement splicing-inactive S100 fractions and alter splice-site selection in alternatively spliced pre-mRNAs. Both

§ Present address: Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138, USA.

proteins function at very early steps in the splicing reaction, being required for assembly of splicing complexes^{26,33}. In certain *in vitro* assays, several of the SR proteins are interchangeable^{32,34}. More recently, however, clear differences in their activities have been observed^{35–38}.

Here we provide evidence that ASF/SF2 can stimulate binding of U1 snRNP to a 5'-splice-site-containing pre-mRNA. This activity requires a region of the RS domain previously shown to be necessary for essential splicing activity³⁹. We also show that ASF/SF2 can bind purified U1 snRNPs in the absence of pre-mRNA, suggesting that binding occurs through protein-protein interactions. Indeed, we show that U1 70K and ASF/SF2 can interact directly, and that this interaction involves the RS domain of each protein. These data suggest that ASF/SF2 aids in the recognition of 5' splice sites by U1 snRNP.

Cooperation of ASF/SF2 and U1 snRNP

Although purified U1 snRNP can interact with 5'-splice-site-containing RNAs¹⁸, this interaction is unlikely to be sufficiently

strong to define the 5' splice site. Because of their ability to influence 5'-splice-site selection, and their requirement early in the splicing reaction, SR proteins are candidates to cooperate with U1 snRNP in 5'-splice-site recognition. We therefore tested whether ASF/SF2 could stabilize the interaction of U1 snRNP with a model pre-mRNA. In Fig. 1a, purified U1 snRNP⁴⁰ and/or recombinant ASF/SF2 (ASF-1) was incubated with ³²P-labelled PIP7.A, a pre-mRNA derived from the adenovirus major late-transcription unit⁴¹, and complexes were resolved by non-denaturing polyacrylamide gel electrophoresis. Only very low levels of complex formation were observed between U1 snRNP alone and PIP7.A (lane 2). The same was true when purified recombinant ASF/SF2 was incubated with PIP7.A (lane 3). But when both U1 snRNP and ASF/SF2 were incubated with the pre-mRNA, two very strong complexes were detected (lane 6, arrows). These complexes migrated in the gel with the nonspecific H complexes⁴² seen when PIP7.A was incubated with HeLa nuclear extract in the absence of added ATP (lane 1), and have a mobility similar to the faint complexes detected with

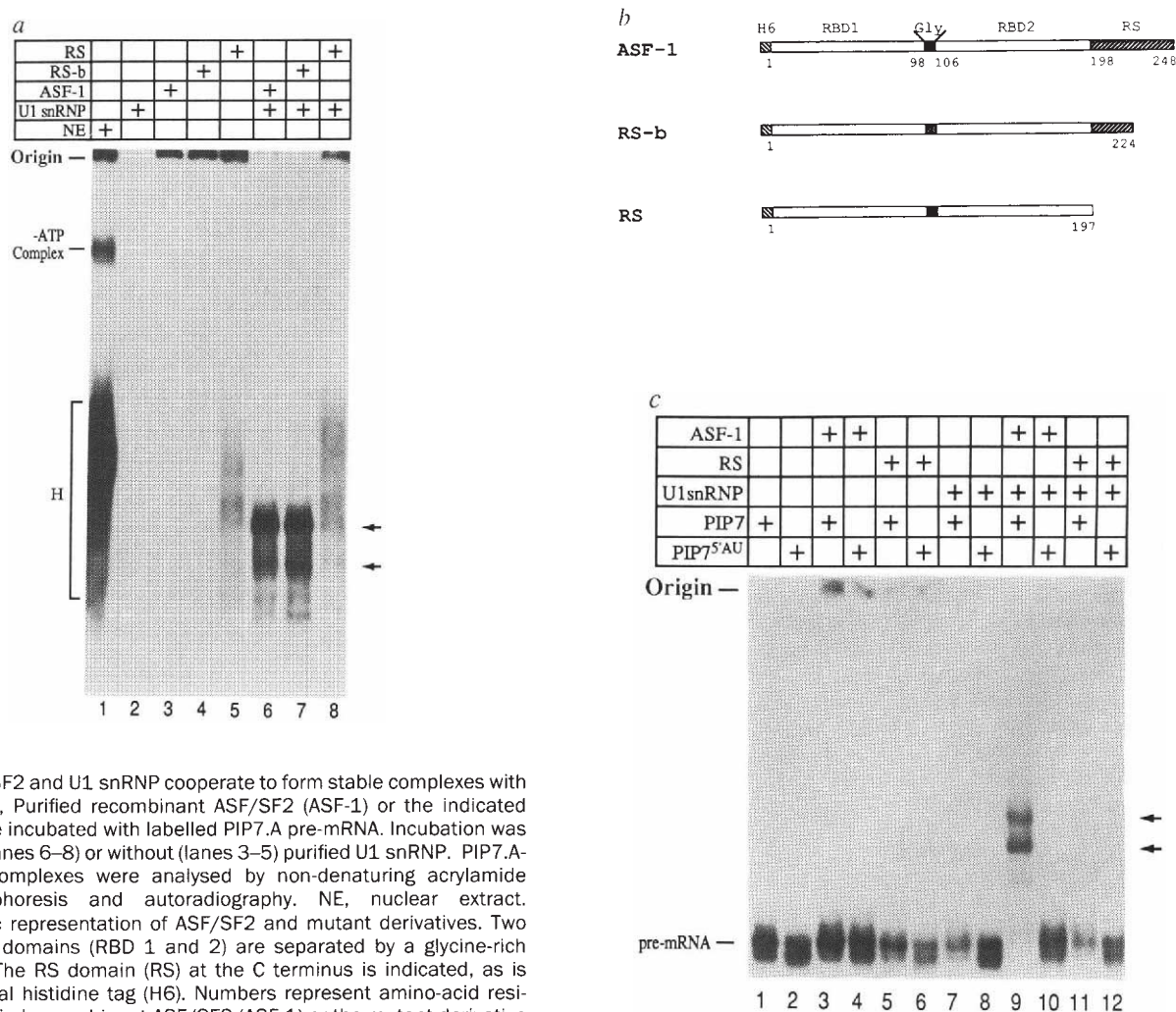


FIG. 1 ASF/SF2 and U1 snRNP cooperate to form stable complexes with pre-mRNA. **a**, Purified recombinant ASF/SF2 (ASF-1) or the indicated variants were incubated with labelled PIP7.A pre-mRNA. Incubation was either with (lanes 6–8) or without (lanes 3–5) purified U1 snRNP. PIP7.A-containing complexes were analysed by non-denaturing acrylamide gel electrophoresis and autoradiography. NE, nuclear extract. **b**, Schematic representation of ASF/SF2 and mutant derivatives. Two RNA-binding domains (RBD 1 and 2) are separated by a glycine-rich hinge (Gly). The RS domain (RS) at the C terminus is indicated, as is the N-terminal histidine tag (H6). Numbers represent amino-acid residues. **c**, Purified recombinant ASF/SF2 (ASF-1) or the mutant derivative RS was incubated with PIP7.A (odd lanes) or the 5'-splice-site mutant PIP7^{5AU} (even lanes). The arrows on the right on both **a** and **c** indicate the positions of the two strong complexes formed only upon incubation of pre-mRNA with ASF/SF2 and U1 snRNP.

METHODS. PIP7.A and PIP7^{5AU} pre-mRNAs were synthesized and uniformly labelled as before⁴¹ and U1 snRNP was purified as described⁴⁰. Recombinant His-tagged ASF/SF2 and mutant derivatives³⁹ were purified from bacterial lysates by Ni²⁺ chromatography²². ASF/SF2 proteins (240 ng; 7.3 pmol), pre-mRNA

(16 ng; 2.4 fmol), and U1 snRNP (800 ng; 2.7 pmol) were incubated under splicing conditions for 5 min at 30 °C, except that no ATP or creatine phosphate were added. Heparin was then added to a final concentration of 0.5 mg ml⁻¹ and reaction mixtures were incubated at 30 °C for an additional 5 min. Samples were loaded on non-denaturing 4% acrylamide:bisacrylamide (80:1) gels containing 50 mM Tris–50 mM glycine⁴¹. Electrophoresis was at 14 V cm⁻¹ for 3 h.

U1 snRNP alone (lane 2). The faster-migrating of the two complexes may reflect loss of a protein component from U1 snRNP. Supporting this view, when a preparation of U1 snRNP lacking the U1-specific A protein (ref. 18; see also below) was used in this assay, only the faster-migrating species was detected (results not shown).

Evidence that the ASF/SF2-U1 snRNP-pre-mRNA interaction is functionally significant was obtained using mutants of ASF/SF2 (Fig. 1b) that vary in their ability to function as essential splicing factors³⁹. ASF/SF2 and RS-b, a mutant lacking the C-terminal 24 amino acids, which is half of the RS domain, are both fully functional in activating a splicing-deficient HeLa S100 extract. When mixed with U1 snRNP, RS-b was as efficient as ASF/SF2 in complex formation with PIP7.A RNA (Fig. 1a, lane 7). In contrast, an inactive mutant derivative of ASF/SF2 lacking the entire RS domain, RS, was unable to form complexes when incubated with U1 snRNP and PIP7.A (Fig. 1a, lane 8). These findings suggest that the ability of ASF/SF2 to cooperate with U1 snRNP in pre-mRNA binding is relevant to the role of the protein in activating splicing, as both processes require a functional RS domain.

To characterize further the specificity of the complexes and to test directly their role in 5' splice site recognition, we analysed a mutant version of PIP7.A pre-mRNA with a single point mutation in the first base of the intron, PIP7^{5'AU} (Fig. 1c), which is

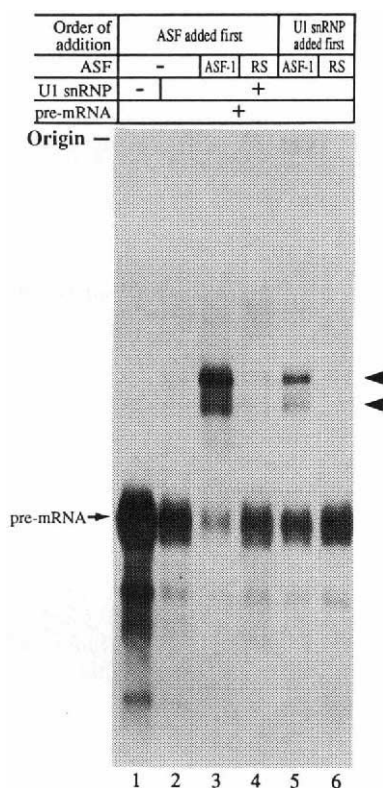


FIG. 2 ASF/SF2 interacts with pre-mRNA first and recruits U1 snRNP. Purified recombinant ASF/SF2 (ASF1) or the RS mutant was incubated with ³²P-labelled PIP7.A pre-mRNA and U1 snRNP as indicated. Preincubations were set up in which either ASF/SF2 or RS was added first to PIP7.A pre-mRNA and incubated at 4 °C for 2 min (lanes 3 and 4), or U1 snRNP was added first to the pre-mRNA at 4 °C and incubated for 2 min (lanes 5 and 6). U1 snRNP (lanes 3 and 4) or ASF/SF2 or RS (lanes 5 and 6) was then added, and all reaction mixtures were incubated at 30 °C for 5 min, followed by a 5-min heparin chase as in Fig. 1. PIP7.A-containing complexes were analysed by non-denaturing polyacrylamide gel electrophoresis and autoradiography as in Fig. 1, except the gel was run for a shorter time in order to observe free PIP7.A RNA (pre-mRNA).

identical to PIP7.A except that the invariant GU dinucleotide at the 5' splice site is changed to AU. Whereas PIP7.A again gave rise to very low levels of complexes with U1 snRNP alone, PIP7^{5'AU} did not (lane 7 and 8). But when it was incubated with ASF/SF2 and U1 snRNP, PIP7^{5'AU} was unable to form complexes (lane 10), whereas PIP7.A gave the same complexes (lane 9, arrows) as shown in Fig. 1a (lane 6). The ASF/SF2 RS mutant was again unable to form complexes (lanes 11 and 12). As above, no complexes were detected in the absence of U1 snRNP (lanes 1–6). Results (not shown) were nearly identical with the related PIP9 pre-mRNA and a corresponding 5' splice site mutant. These data indicate that an intact, functional

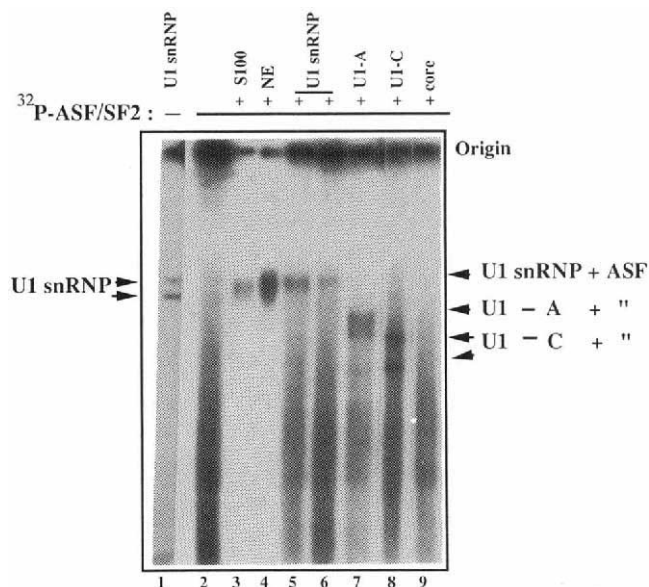


FIG. 3 ASF/SF2 can bind U1 snRNPs in the absence of pre-mRNA. ³²P-labelled ASF/SF2 and the indicated U1 snRNP preparations were incubated and complexes detected by non-denaturing electrophoresis and autoradiography. The migration of U1 snRNP in the absence of ASF/SF2 (lane 1) is indicated by the arrows on the left. ASF/SF2 alone is shown in lane 2. Complexes formed when ASF/SF2 was incubated with HeLa S100 (lane 3) or nuclear extract (lane 4) are shown. Complexes formed when labelled ASF/SF2 was incubated with either purified U1 snRNP (lanes 5 and 6) or U1 snRNP lacking A and C proteins (lanes 7 and 8, respectively) are indicated by the arrows on the right. No complexes were formed when core U1 snRNP were used.

METHODS. Recombinant ASF/SF2 was phosphorylated by protein kinase C (Sigma) in the presence of [γ -³²P]ATP. The kinase reaction mixture contained 3 μ g of recombinant, purified ASF/SF2, 12 units protein kinase C, 50 mM potassium phosphate, pH 6.8, 50 mM MgSO₄, 330 μ Ci [γ -³²P]ATP (>7,000 Ci per mmol; ICN), 0.2 mM PMSF. Unincorporated nucleotides were removed by spun column chromatography (1 ml) through Sephadex G-25 which had been pre-equilibrated in modified buffer D²¹. Specific activities of 1–3 $\times 10^7$ c.p.m. μ g⁻¹ were obtained. Binding reactions (12.5 μ l) contained 600ng (lane 7) or 300ng (lane 8) purified U1 snRNP or 600ng U1 snRNPs depleted of A (U1-A), C (U1-C) or 70K, A and C (core)¹⁸, 10⁶ c.p.m. ³²P-ASF/SF2, 2 mM MgCl₂, 20 mM KCl, 1 mM ATP, 5 mM creatine phosphate, 1.2% polyvinylalcohol, 4 units RNasin, 4% glycerol, 0.1 mM DTT, and protease inhibitor cocktail (0.2 mM PMSF, 2 μ g ml⁻¹ leupeptin, 2 μ g ml⁻¹ pepstatin). The purified U1 snRNP shown in lane 1 was labelled with its endogenous kinase¹⁹. Similarly labelled -A and -C particles displayed mobilities slightly greater than the corresponding complexes detected with ASF/SF2 (results not shown). In reactions containing HeLa S100 or nuclear extracts (prepared as described²¹), 4 μ l extract was substituted for the purified U1 snRNP. Binding was done at 30 °C for 30 min, and reaction mixtures were loaded on a 6% polyacrylamide gel (30:1), 0.5 $\times$ TBE, and electrophoresed at 3.5 watts until the xylene cyanol dye had migrated 8.5 cm. The gel was dried and autoradiographed. Lane 1 was exposed for 3 days, the rest of the gel for 1 day.

5' splice site is required for formation of the U1 snRNP-ASF/SF2-RNA complex.

ASF/SF2 binds first

Our results argue in favour of the possibility that ASF/SF2 and U1 snRNP are two of the earliest components in the formation of splicing complexes leading to the assembly of the spliceosome. To test whether it is important that U1 snRNP or ASF/SF2 interact with the pre-mRNA first, we repeated the experiments

shown in Fig. 1 but altering the order of addition of ASF/SF2 and U1 snRNP. When ASF/SF2 was added first to the pre-mRNA, as before, complex formation was efficient (Fig. 2, lane 3). In contrast, when U1 snRNP was added to the pre-mRNA first, complexes were detected at much lower levels (lane 5). These data indicate that there is a preferred order in the formation of these complexes, in which ASF/SF2 interacts first with the pre-mRNA and then recruits U1 snRNP to the 5' splice site. This is consistent with experiments showing that pre-binding to SR proteins gives a kinetic advantage to pre-mRNAs in subsequent splicing reactions³⁸.

ASF/SF2 can bind U1 snRNP

Our results are consistent with the idea that ASF/SF2 and U1 snRNP can interact directly. To test this, we asked whether an interaction between the two factors could occur in the absence of pre-mRNA. We labelled ASF/SF2 with [γ -³²P]ATP, repurified it, incubated with purified U1 snRNPs, and visualized the resulting complexes after non-denaturing polyacrylamide gel electrophoresis (Fig. 3 legend). Figure 3 shows that ³²P-ASF/SF2 formed stable complexes with purified U1 snRNP (lanes 5 and 6). Strikingly, this complex co-migrated with complexes formed when HeLa S100 (lane 3) or nuclear extract (lane 4) was used instead of purified U1 snRNP, and migrated slightly slower than ³²P-labelled U1 snRNP electrophoresed in the absence of ASF/SF2 (lane 1). The specificity of the interaction is suggested by the fact that only a single complex was observed when S100 or nuclear extracts were incubated with ASF/SF2. Using conditions previously described¹⁸, we prepared U1 snRNPs lacking either the types A or C U1-specific proteins. Both preparations (lanes 7 and 8) also formed complexes with ASF/SF2, indicating that neither of these proteins is necessary for the interaction. In contrast, no complexes were detected with a U1 snRNP derivative lacking all three U1-specific proteins, 70K, A and C ('core', lane 9). These experiments show that ASF/SF2 can interact with U1 snRNP in the absence of pre-mRNA, and suggest that the interaction may involve the U1 70K protein.

ASF/SF2 and U1 70K protein interact directly

The sequence motifs present in the U1 70K protein⁴³ suggest a similarity with ASF/SF2: both contain N-terminal RBDs and C-terminal RS domains. To investigate whether the ASF/SF2-U1 snRNP interaction might be mediated by protein-protein interactions involving U1 70K, we first used far-western blotting (Fig. 4). For this, full-length U1 70K (WT) and a truncated derivative lacking the RS region (Δ RS) were synthesized by *in vitro* transcription and translation, and incubated with filters containing recombinant ASF/SF2 that had been resolved by SDS-polyacrylamide gel electrophoresis (Fig. 4, far western). Strikingly, U1 70K WT bound ASF/SF2 efficiently, whereas no interaction was detected with Δ RS. Binding was completely resistant to treatment with RNase A (results not shown), indicating that RNA was not an intermediary in the interaction. Also, Mg²⁺, which can cause aggregation of SR proteins³², was not required for binding (results not shown). Both filters were washed and incubated with anti-ASF/SF2 antibodies to verify that they contained equal amounts of ASF/SF2 (western). The preparation of recombinant ASF/SF2 used for Fig. 4 consisted of two species, of which only the larger was able to interact with U1 70K. The smaller form was a proteolytic product lacking all or most of the RS domain (unpublished data), and its failure to bind U1 70K suggests that the RS domain of ASF/SF2 is essential for binding.

To delineate the regions of ASF/SF2 required for interaction with U1 70K in more detail, a number of deletion mutants were tested by far-western analysis (Fig. 5a). Of these, the mutant lacking the RS domain (lane 7) showed significantly decreased binding compared with full-length ASF/SF2 (lane 8), confirming that the RS domain of ASF/SF2 is required for efficient binding to U1 70K. But analysis of mutants spanning most of

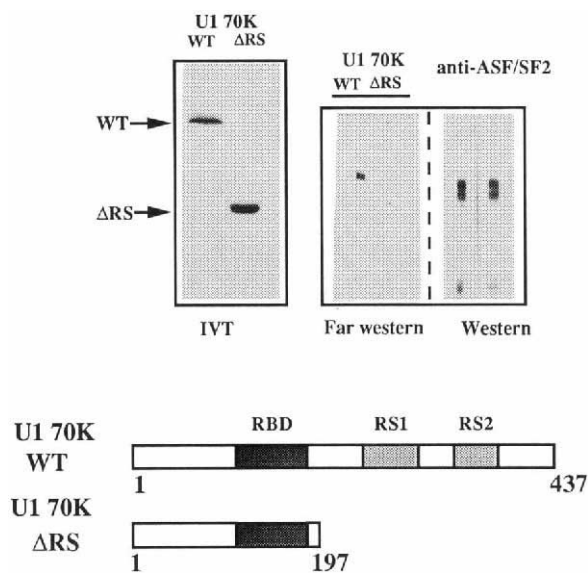


FIG. 4 The U1 snRNP protein and ASF/SF2 interact directly via their RS domains. Full-length U1 70K protein (WT) and a truncated derivative lacking the RS domains (Δ RS) were translated *in vitro* in the presence of ³⁵S-methionine and visualized by SDS-PAGE (IVT, for *in vitro* translated). WT and Δ RS were then used to probe strips of nitrocellulose containing recombinant ASF/SF2 (far western). That equivalent amounts of ASF/SF2 were immobilized in both strips is shown by reprobing the same strips with anti-ASF antibodies (western). Diagrams of U1 70K WT and Δ RS indicating the positions of the RNA-binding domain (RBD) and RS domains (RS 1 and 2) are shown below. Note that the Δ RS derivative used here extends to residue 197. Identical results were obtained with a mutant extending to residue 210 (results not shown). METHODS. A plasmid containing the full-length human U1 70K cDNA⁴³ was cleaved with either *Bam*H1 (WT) or *Kpn*I (Δ RS) to generate templates for *in vitro* transcription and translation. Transcription and translation were performed with T7 RNA polymerase and translation with a rabbit reticulocyte translation kit (Promega) in the presence of ³⁵S-methionine according to manufacturer's instructions. Proteins were separated from unincorporated ³⁵S-methionine by spun column chromatography in a syringe containing 1 ml Sephadex G-25 pre-equilibrated in buffer D. WT U1 70K (2×10^5 c.p.m. ml⁻¹) or U1 70K Δ RS (5×10^5 c.p.m. ml⁻¹) were used as probes for far-western analysis. Recombinant ASF/SF2 (1 μ g per cm strip) was run on 12.5% SDS-PAGE and transferred to nitrocellulose using a 12.5 mM CHES, pH 9.5/20% methanol transfer buffer. Immediately before use, filters were incubated in modified buffer D²¹ containing 6M guanidine HCl for 2 $\times$ 20 min at room temperature. Filters were then rinsed twice in buffer D, and blocked in buffer D containing 5% non-fat dry milk, 2% bovine serum albumin for 1 h, then incubated with labelled proteins in binding buffer (13 mM HEPES, pH 7.9, 28 mM (NH₄)₂SO₄, 33 mM KCl, 3.3 mM MgCl₂, 0.13 mM EDTA, 2% bovine serum albumin, 0.5% non-fat dry milk, 0.1% Tween 20, 0.5 mM DTT, 0.5 mM PMSF) for 12–16 h. Blots were washed for 3 $\times$ 15 min in binding buffer without bovine serum albumin and non-fat dry milk, dried and autoradiographed for 2–3 d. The same strips were then blocked in 10% non-fat dry milk for 1 h, incubated with a 1:5,000 dilution of a rabbit polyclonal serum raised against recombinant ASF/SF2 for 2 h, washed 3 $\times$ 5 min, incubated with a 0.1 μ Ci ml⁻¹ (1 ml per strip) goat anti-rabbit ¹²⁵I-labelled antibody (NEN) for 1 h, washed 3 $\times$ 10 min, and autoradiographed at -70 °C for 24 h.

the ASF/SF2 protein (lanes 1–6) failed to detect another region of the protein essential for the ASF/SF2–U1(70K) interaction, confirming that the ASF/SF2 RS domain is critical for binding.

ASF/SF2 belongs to the SR family of proteins, the members of which (SR proteins p75, p55, p40, p30a (SC35), p30b (ASF/SF2) and p20) can be purified by a two-step procedure from either HeLa or calf thymus extracts³². Our findings raise the question of whether other SR proteins apart from ASF/SF2 can interact with U1 70K. We therefore analysed a preparation of calf thymus SR proteins by far-western blotting using U1 70K translated *in vitro* as a probe (Fig. 5b). A duplicate blot was

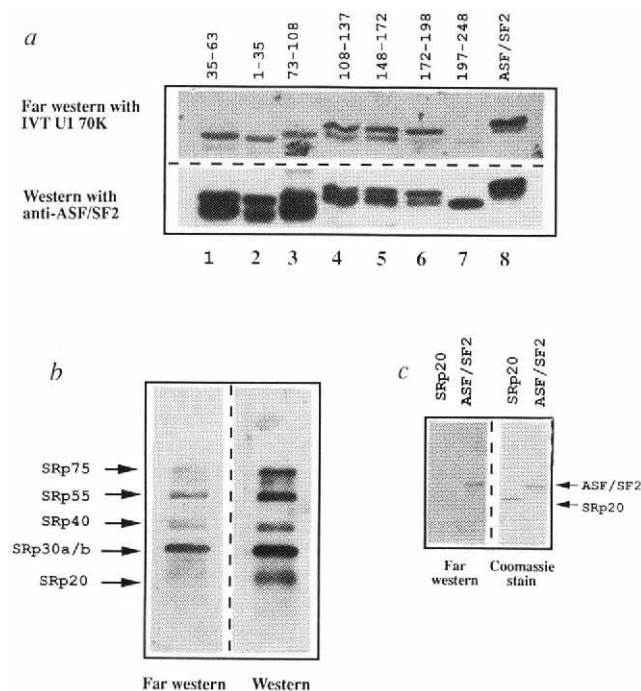


FIG. 5 The ASF/SF2 RS domain is required for interaction with U1 70K. The far-western method was used to determine the domains on ASF/SF2 required for interaction with U1 70K. Full-length, *in vitro* translated U1 70K was used as a probe as described in Fig. 4. Deletion mutants of ASF/SF2 were expressed in *E. coli* as described³⁹. a, Full-length ASF/SF2 and variants were resolved by SDS-PAGE, transferred to nitrocellulose, probed with U1 70K protein, and autoradiographed. Amino-acid residues deleted from ASF/SF2 are indicated above lanes 1–7. Lane 8 contains full-length ASF/SF2. Protein blots were re-probed with anti-ASF/SF2 antibodies as described in the legend to Fig. 4. b, Calf thymus SR proteins were resolved by SDS-PAGE, transferred to nitrocellulose, and probed with U1 70K (far western). A duplicate blot was probed with mAb 104 (western). Arrows indicate the migration of SRp75, SRp55, SRp40, SRp30a/b and SRp20. c, Recombinant SRp20 and ASF/SF2 were resolved by SDS-PAGE, transferred to nitrocellulose and probed with U1 70K (far western). Equivalent amounts of SRp20 and ASF/SF2 were stained with Coomassie blue (shown on the right). Arrows indicate the positions of ASF/SF2 and SRp20.

probed with a monoclonal antibody (mAb 104) against a phospho-epitope shared by all SR proteins³² (Fig. 5b, western blot). The results show that one or both SRp30 proteins (namely ASF/SF2 and SC35), which are not resolved by this method, bound U1 70K, as did SRp55 and SRp40. No interaction with SRp20 was detected, and binding to SRp75 was weak. The lack of interaction with SRp20 was confirmed by probing recombinant SRp20 (ref. 44), which also failed to bind U1 70K (Fig. 5c). U1 70K can thus interact with some but not all SR proteins.

To confirm the ASF/SF2-U1 70K interaction by an independent method, we asked whether the two proteins produced by *in*

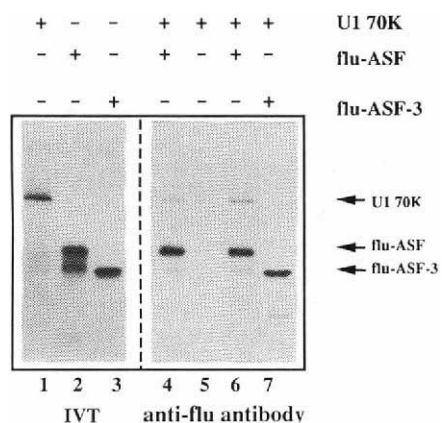


FIG. 6 U1 70K and ASF/SF2 can be co-immunoprecipitated. Complexes of U1 70K and ASF/SF2 were immunoprecipitated using a 'flu-tagged ASF/SF2 fusion protein (flu-ASF) and anti-flu antibodies, and separated by SDS-PAGE. The indicated *in vitro* translated proteins before immunoprecipitation are shown on the left (IVT). Combinations of proteins were mixed and precipitated by anti-flu antibodies as follows: lane 4, U1 70K and flu-ASF (co-translated); lane 5, U1 70K alone; lane 6, U1 70K and flu-ASF (mixed after translation); and lane 7; U1 70K and flu-ASF-3.

METHODS. An ASF/SF2 cDNA construct containing sequences encoding a 21-residue influenza haemagglutinin epitope⁴⁵ at the 5' end (flu-ASF) was subcloned into pBluescript (SK+), cleaved with XbaI, transcribed and *in vitro* translated as described in Fig. 4 legend. The same was done with a cDNA encoding ASF-3 (flu-ASF-3). Full-length U1 70K was also produced by *in vitro* translation. Aliquots (4 µl) were then mixed and immunoprecipitated with the anti-flu monoclonal antibody 12CA5 (ref. 45) using the following protocol: 100 µl of 12CA5 culture supernatant was incubated with 10 µl of protein A-agarose overnight at 4 °C, washed twice in NET/2 (0.15 M NaCl, 50 mM Tris, pH 7.4, 0.05% NP-40) and once in IPBB (13 mM HEPES, pH 7.9, 28 mM (NH₄)₂SO₄, 3.3 mM MgCl₂, 100 mM KCl, 0.1% Tween 20, 0.5 mM DTT, 0.2 mM PMSF, 2 µg ml⁻¹ leupeptin, and 2 µg ml⁻¹ pepstatin). Antibody-bound protein A-agarose was then incubated with *in vitro*-translated protein in 400 µl IPBB containing 2% protease-free bovine serum albumin for 4 h at 4 °C. Protein A-agarose-bound complexes were centrifuged, washed three times in IPBB, three times in IPBB + 0.2 M KCl, twice in IPBB containing 0.4% Tween 80 + 0.2 M KCl, resuspended in SDS-PAGE sample buffer, boiled and loaded onto a 12.5% SDS-polyacrylamide gel. Before drying and autoradiography, the gel was soaked in 10% acetic acid / 25% isopropanol for 30 min, rinsed 3 × in H₂O, and incubated in 1 M salicylic acid for 1 h.

vitro translation could be co-immunoprecipitated. We therefore synthesized derivatives of ASF/SF2 and ASF-3 (a form arising by alternative splicing and lacking the RS domain²¹ as fusion proteins containing an N-terminal influenza virus haemagglutinin epitope⁴⁵), incubated them with U1 70K and immunoprecipitated with the monoclonal antibody 12CA5, which recognizes the 'flu epitope. Figure 6 shows that U1 70K was co-immunoprecipitated with 'flu-ASF/SF2, but not with 'flu-ASF-3. Formation of the ASF/SF2-U1(70K) complex did not require co-translation (compare lane 4 (co-translated) and 6 (mixed after translation)). These results are additional evidence that ASF/SF2 and U1 70K can form a stable complex, and that the ASF/SF2 RS domain is required for this interaction.

Discussion

We have shown that ASF/SF2 can greatly enhance the binding of U1 snRNP to pre-mRNA, and that this interaction is probably mediated by specific protein-protein interactions involving the RS domain of ASF/SF2 and a similar region of U1 70K. The complex formed by ASF/SF2, U1 snRNP and pre-mRNA is exceptionally stable (our unpublished results) and probably involves cooperative binding of these factors. The size of the complex is close to that of the commitment complex¹⁴. We envisage that the ASF/SF2-U1 snRNP-pre-mRNA interaction defines the 5' splice site, but is not sufficient to define the 3' splice site. This probably requires other factors such as U2AF. Indeed, our finding that ASF/SF2 facilitates U1 snRNP binding to the 5' splice site suggests a functional similarity with U2AF, which enhances binding of U2 snRNP to the branch site^{27,28}. The U2AF RS domain is necessary for splicing activation, but not for sequence-specific RNA binding²⁹. Whether it, like the ASF/SF2 RS domain, also participates in protein-protein interactions is not known.

The data presented here, coupled with other recent results, suggest a model for 5'-splice-site recognition in which ASF/SF2 binds the pre-mRNA and then functions to help recruit U1 snRNP to the 5' splice site. We have shown that purified ASF/SF2 can specifically recognize 5' splice sites in two different pre-mRNAs⁴⁶, and this may reflect a prelude to its interaction with U1 snRNP. Although ASF/SF2 by itself does not efficiently bind the PIP7A pre-mRNA used here, perhaps in this case the

ASF/SF2-RNA interaction is initially unstable, but is then stabilized by interaction with U1 snRNP. This is supported by our order-of-addition experiment indicating that ASF/SF2 must be mixed with the pre-mRNA before U1 snRNP for efficient complex formation. Also pre-incubation of either of two different pre-mRNAs with purified ASF/SF2 or SC35 is sufficient to commit the RNAs to splicing³⁸. Taken together, these findings suggest that ASF/SF2 and probably other SR proteins (perhaps in a substrate-specific manner³⁸) function at the earliest step in spliceosome assembly and do so by stabilizing U1 snRNP-5'-splice-site interactions. This fits nicely with the ability of ASF/SF2 to influence 5'-splice-site selection in alternatively spliced pre-mRNAs²¹⁻²⁴.

RS domains have now been detected in several metazoan splicing factors. These include, in addition to the mammalian factors already described, several splicing regulators in *Drosophila*⁴⁷. RS domains have not been found in *Saccharomyces cerevisiae*, which may be indicative of a mechanistic difference in an early step in splicing complex assembly. Indeed, the apparent yeast homologue of U1 70K specifically lacks an RS region⁴⁸. The precise role of RS domains in splicing has been unclear. Given their high positive charge, one possibility was that they interact with RNA. Although our results do not exclude this, they do indicate that these domains can mediate specific protein-protein interactions. We have shown that the RS domain of ASF/SF2 and a similar region in U1 70K are required for such interaction, but we do not know whether these two highly charged regions contact each other directly. However, no other specific region of ASF/SF2 was found to be essential for binding. Although further investigation is necessary into how the two proteins interact, recent findings that serine residues in both the U1 70K and ASF/SF2 RS domains can be phosphorylated by a U1 snRNP-associated kinase^{19,49} raise the possibility that this interaction may be subject to regulation during splicing. Finally, our results suggest a possible similarity in the mechanisms used by transcription activators (for review, see ref. 50) and splicing activators such as ASF/SF2. In each case, specific protein-protein interactions between the DNA (or RNA)-bound activator and a component of the general transcription (or splicing) machinery facilitate formation of productive transcription (or splicing) complexes. □

Received 1 September 1993; accepted 17 January 1994.

- Green, M. R. *A. Rev. Cell Biol.* **7**, 559-599 (1991).
- Moore, M. J., Query, C. C. & Sharp, P. A. In *RNA World* (eds Gestland, R. & Atkins, J.) 303-357 (Cold Spring Harbor Laboratory Press, New York, 1993).
- Smith, C. W. J., Patton, J. G. & Nadal-Ginard, B. *A. Rev. Genet.* **3**, 527-577 (1989).
- Steitz, J. A. *et al.* in *Structure and Function of Major and Minor snRNPs* (ed. Birnstiel, M. L.) 115-154 (Springer, Berlin, 1988).
- Zhuang, Y. & Weiner, A. M. *Cell* **46**, 827-835 (1986).
- Siliciano, P. G. & Guthrie, C. *Genes Dev.* **2**, 1258-1267 (1988).
- Seraphin, B., Kretzner, L. & Rosbash, M. *EMBO J.* **7**, 2533-2538 (1988).
- Newman, A. & Norman, C. *Cell* **68**, 743-754 (1992).
- Sawa, H. & Shimura, Y. *Genes Dev.* **6**, 244-254 (1992).
- Wasserman, D. A. & Steitz, J. A. *Science* **257**, 1918-1925 (1992).
- Ruby, S. W. & Abelson, J. *Science* **242**, 1028-1035 (1988).
- Seraphin, B. & Rosbash, M. *Cell* **59**, 349-358 (1989).
- Michaud, S. & Reed, R. *Genes Dev.* **5**, 2534-2546 (1991).
- Jamison, S. F., Crow, A. & Garcia-Blanco, M. *Molec. cell. Biol.* **12**, 4279-4287 (1992).
- Mount, S. M. *Nucleic Acids Res.* **10**, 459-472 (1982).
- Mount, S. M., Pettersson, L., Hinterberger, M., Karmas, A. & Steitz, J. A. *Cell* **33**, 509-519 (1983).
- Tatei, K., Takemura, K., Tanaka, H., Masaki, T. & Ohshima, V. *J. Biol. Chem.* **262**, 11667-11673 (1987).
- Heinrichs, V., Bach, M., Winkelmann, G. & Lührmann, R. *Science* **247**, 69-72 (1990).
- Tazi, J. *et al.* *Nature* **363**, 283-286 (1993).
- Chabot, B. & Steitz, J. A. *Molec. cell. Biol.* **7**, 698-707 (1987).
- Ge, H. & Manley, J. L. *Cell* **62**, 25-34 (1990).
- Ge, H., Zuo, P. & Manley, J. L. *Cell* **66**, 373-382 (1991).
- Krainer, A. R., Conway, G. C. & Kozak, D. *Cell* **62**, 35-42 (1990).
- Krainer, A. R., Mayeda, A., Kozak, D. & Binns, G. *Cell* **66**, 383-394 (1991).
- Fu, X. D. & Maniatis, T. *Science* **256**, 535-538 (1992).
- Fu, X. D. & Maniatis, T. *Proc. natn. Acad. Sci. U.S.A.* **89**, 1725-1729 (1992).
- Ruskin, B., Zamore, P. D. & Green, M. R. *Cell* **52**, 207-219 (1988).
- Zamore, P. D. & Green, M. R. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9243-9247 (1989).

- Zamore, P. D., Patton, J. G. & Green, M. R. *Nature* **355**, 609-614 (1992).
- Bandziulis, R. J., Swanson, M. S. & Dreyfuss, G. *Genes Dev.* **3**, 431-437 (1989).
- Kenan, D. J., Query, C. C. & Keene, J. D. *Trends biochem. Sci.* **16**, 214-220 (1991).
- Zahler, A. M., Lone, W. S., Stalk, J. A. & Roth, M. B. *Genes Dev.* 837-847 (1992).
- Krainer, A. R., Conway, G. C. & Kozak, D. *Genes Dev.* **4**, 1158-1171 (1990).
- Fu, X. D., Mayeda, A., Maniatis, T. & Krainer, A. *Proc. natn. Acad. Sci. U.S.A.* **89**, 11224-11228 (1992).
- Kim, Y.-J., Zuo, P., Manley, J. L. & Baker, B. S. *Genes Dev.* **6**, 2569-2579 (1992).
- Zahler, A. M., Neugebauer, K. M., Lane, W. S. & Roth, M. B. *Science* **260**, 219-222 (1993).
- Tian, M. & Maniatis, T. *Cell* **74**, 105-114 (1993).
- Fu, X. D. *Nature* **365**, 82-85 (1993).
- Zuo, P. & Manley, J. L. *EMBO J.* **12**, 4727-4737 (1993).
- Bach, M., Bringmann, P. & Lührmann, R. *Meth. Enzym.* **181**, 232-256 (1990).
- Gil, A., Sharp, P. A., Jamison, S. F. & Garcia-Blanco, M. A. *Genes Dev.* **5**, 1224-1236 (1991).
- Konarska, M. M. & Sharp, P. A. *Cell* **46**, 845-855 (1986).
- Query, C. C., Bentley, R. C. & Keene, J. D. *Cell* **57**, 89-101 (1989).
- Ayane, M., Preuss, U., Kohler, G. & Nielsen, P. J. *Nucleic Acids Res.* **19**, 1273-1278 (1991).
- Field, J. *et al.* *Molec. cell. Biol.* **8**, 2159-2165 (1988).
- Zuo, P. & Manley, J. L. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Baker, B. S. *Nature* **34**, 521-524 (1989).
- Smith, V. & Barrell, B. *EMBO J.* **10**, 2627-2634 (1991).
- Wopmann, A. *et al.* *Nucleic Acids Res.* **21**, 2815-2822 (1993).
- Ptashne, M. & Gann, A. *Nature* **346**, 329-331 (1990).

ACKNOWLEDGEMENTS. J.D.K. and S.F.J. contributed equally to this work. We thank A. Zahler and M. Roth for calf thymus SR proteins and mAb 104, J. Field for the 12CA5 antibody, R. Tacke for recombinant SRp20, J. Keene for U1 70K cDNA, and J. Harper, Y. Takagaki, R. Tacke, R. Mancebo, S. Mount and S. Koltz for discussion. This work was supported by a grant from the NIH to J.L.M., from the ACS to M.G.B., and from DFG to R.L.

The rarity of soft γ -ray repeaters deduced from reactivation of SGR1806 – 20

C. Kouveliotou^{*†}, G. J. Fishman^{*}, C. A. Meegan^{*},
W. S. Paciesas[‡], J. van Paradijs[§], J. P. Norris^{||},
R. D. Preece^{*‡}, M. S. Briggs[‡], J. M. Horack^{*},
G. N. Pendleton[‡] & D. A. Green[¶]

^{*} Space Science Laboratory, ES66,

NASA/Marshall Space Flight Center, Huntsville, Alabama 35812, USA

[†] Universities Space Research Association, 4950 Corporate Drive,
Suite 100, Huntsville, Alabama 35806, USA

[‡] Department of Physics, University of Alabama in Huntsville,
Alabama 35899, USA

[§] Astronomical Institute "Anton Pannekoek"

& Centre for High-Energy Astrophysics, Kruislaan 403,
1098 SJ Amsterdam, The Netherlands

^{||} Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

[¶] Mullard Radio Astronomy Observatory, Cavendish Laboratory,
Madingley Road, Cambridge CB3 0HE, UK

ONLY two different types of γ -ray transient sources are presently known: over one thousand γ -ray bursters (GRBs) and only three soft γ -ray repeaters (SGRs). The latter are distinguished by their propensity for recurrent burst behaviour^{1–3}, in contrast to the non-repeating GRB sources. Recurrent emission from one of the repeaters, SGR1900 + 14, has been detected⁴ earlier by the Burst and Transient Source Experiment (BATSE) aboard the Compton Gamma-Ray Observatory. Here we report renewed burst activity from SGR1806 – 20, the most prolific of the three known SGRs. This detection of reactivation of this source has been rapidly followed by identification of an X-ray counterpart^{5,6}, which also coincides with a compact radio source⁷ now identified as a plerionic (pulsar-powered) supernova remnant⁸. In combination, these results are leading to a convergence of ideas about the nature of SGRs, which can now be firmly identified as neutron stars. That BATSE has detected no new sources in its two and a half years of operation indicates that SGRs are rare in our Galaxy.

SGR1806 – 20 is distinguished from the other two members of its class as the most prolific emitter: between 1979 and 1985, it emitted over 100 bursts⁹, compared to the 16 and 3 emissions of the other two sources (SGR0526 – 66 and SGR1900 + 14, respectively)⁹. Its outbursts were clustered stochastically with separations covering all timescales, from seconds to years². On 29 September 1993, BATSE detected 3 bursts within 14 hours; the source appeared again once on 5 Oct, 9 Oct and 10 Nov, respectively. Several weaker emissions, possibly related to the source, were also located within 20° of SGR1806 – 20 during a background search covering the period 29 September to 9 October; these bursts were not intense enough to trigger the instrument. An exhaustive search covering several weeks of data, before and after this interval, is currently underway.

Figure 1 depicts the time profiles of all six triggered bursts from SGR1806 – 20 at the same timescale. Four of the six bursts (*a*, *d*–*f*) have simple, triangular profiles. We have fitted the high-time-resolution data (binned as necessary with 1, 2, or 4 ms) with a model pulse¹⁰ and calculated the duration encompassing 90% of the total counts, similar to the T_{90} described previously¹¹. The durations vary from 30 ms to ~1 s; the rise times are consistently shorter than the decay times (Table 1). In all cases, we see no pulsations during the events in the millisecond range.

The third event of 29 September 1993, merits particular attention. Figure 1*c* (lower panel) shows the 1993 BATSE burst; Fig. 1*c* (upper panel) shows another event from the same source detected³ with the Solar Maximum Mission (SMM) in November 1983. The timescales and the temporal resolutions are the same (0.128 ms). There is a striking similarity in the temporal

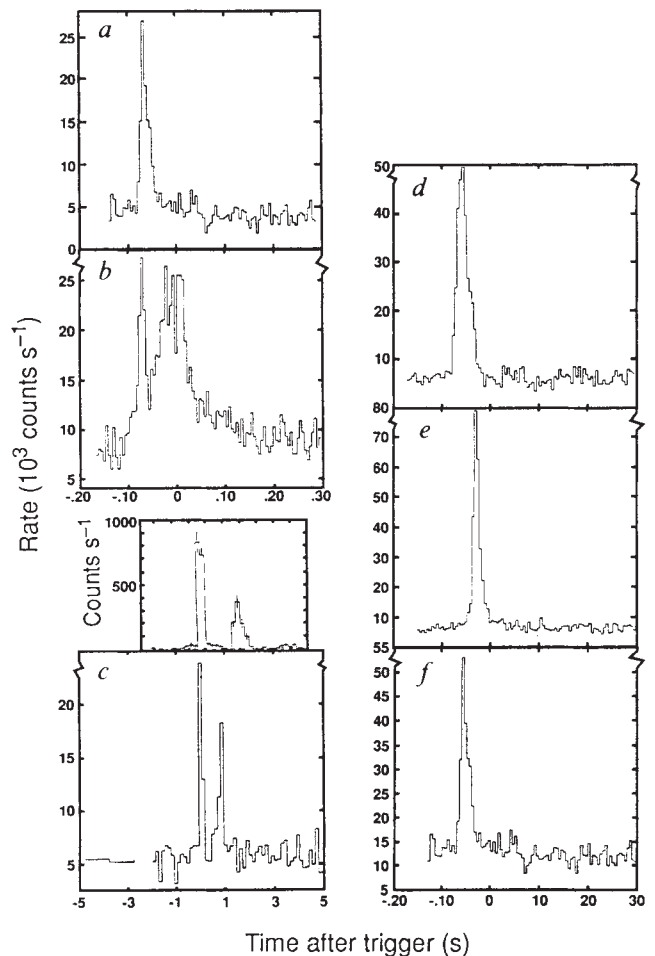


FIG. 1 Light curves of the six SGR bursts detected with BATSE on the Compton Gamma Ray Observatory (CGRO). The counts are integrated between 25 and 130 keV. All timescales are 500 ms total duration with 5 ms resolution. *a*, First event of 29 Sept 1993. *b*, Second event of 29 Sept, 1993. *c*, Two emissions of SGR1806 – 20 separated by almost 10 yr: timescales and temporal resolutions are the same (0.128 s). Upper panel, event detected with Solar Maximum Mission on 16 Nov 1983. The counts are integrated between 32 and 148 keV. Lower panel, event detected with BATSE on 29 Sept, 1993. The counts are integrated between 25 and 130 keV. *d*, Event of 5 Oct, 1993. *e*, Event of 9 Oct, 1993. *f*, Event of 10 Nov, 1993.

characteristics of the two events: they both comprise two pulses ~1 s apart, with the first and more intense pulse lasting ~0.25 s and the second pulse lasting ~0.5 s. The fluence (time-integrated energy flux) of the 1993 event is almost 10 times smaller than that of the 1983 emission. These are the only 2 out of over 100 events from this source that exhibit this compound structure. This structure preservation suggests that the 1 s timescale could be related to the source characteristics.

We have sufficient data for detailed spectral analysis only for the third burst of 29 September 1993. A fit with an optically-thin thermal bremsstrahlung function yields a value of $kT = 28 \pm 5$ keV and a normalization of 1.8 ± 0.6 photons s^{-1} keV^{-1} cm^{-2} at 1 keV (reduced $\chi^2 = 1.3$ for 11 degrees of freedom). The kT value is similar to the ones reported before^{1,3,12,13}, indicating very little change in the emission mechanism over 10 years. The fluences and hardness ratios, R_{21} (ratios of the counts between 50–100 keV and 25–50 keV) cluster around an average value (Table 1), consistent with the assumption that all six outbursts have similar energetics. The peak fluxes of the emissions range between 2 to 9×10^{-6} erg cm^{-2} s^{-1} . These values are 4 to 15

TABLE 1 Summary of SGR1806–20 temporal, spatial and spectral properties

Date	Time (UT)	(RA, dec.)*	Fluence† (erg cm ⁻²)	Rise‡§ (ms)	Decay‡ (ms)	Duration‡ (ms)	R ₂₁ ¶
29 Sept 1993	09:40:01.625	18 h 20 min, –19°	1.2×10^{-7}	7	13	112	0.51 ± 0.09
29 Sept 1993	20:26:21.722	18 h 20 min, –18°	2.8×10^{-7}	*	*	192	0.34 ± 0.02
29 Sept 1993	23:15:32.890	18 h 08 min, –18°	1.4×10^{-7}	...	...	1,024	0.48 ± 0.03
				6 [☆]	6 [☆]	29 [☆]	0.31 ± 0.04
5 Oct 1993	22:51:28.602	18 h 00 min, –18°	1.6×10^{-7}	9	18	50	0.46 ± 0.03
9 Oct 1993	22:24:58.074	18 h 12 min, –26°	1.4×10^{-7}	3	6	28	0.46 ± 0.03
10 Nov 1993	10:49:03.066	17 h 56 min, –25°	9.5×10^{-8}	5	10	40	0.70 ± 0.15

* Right ascension, declination (epoch 2000).

† In the energy range of 25–130 keV.

‡ Calculated between 25 and 50 keV because of better statistics.

§ Time between half-width half-maximum of the pulse before the peak, and the peak.

|| Time between half-width half-maximum of the pulse after the peak, and the peak.

¶ Ratios of the number of counts between 50–100 keV, and 25–50 keV.

* Multi-peaked structure.

☆ First pulse only.

times smaller than those of the brightest events detected from this source⁹ and comparable to the ones detected³ with SMM.

Figure 2 shows the overlapping error circles of the 6 bursts, drawn with the 1σ radius of 8° . The right ascension and declination of the circle centres are given in Table 1. The $1.3^\circ \times 4'$ error box indicated by a short vertical line is the updated (K. Hurley, personal communication) SGR1806–20 location (epoch 2000) calculated from data from the 2nd Inter-Planetary Network (IPN)²³. Clearly, the BATSE circles overlap the IPN location.

Based on the identification of all three emissions on 29 September with SGR1806–20, we initiated an international multi-wavelength campaign to search for counterparts in the IPN SGR error box. As a result, on 9–10 October 1993, the Japanese X-ray satellite ASCA detected² a burst from the area of the source simultaneous with the burst detected with BATSE on 9 October 1993. ASCA also discovered⁵ the steady X-ray counterpart of SGR1806–20.

Earlier searches⁷ had suggested the identification of SGR1806–20 with a supernova remnant (SNR), which was later identified¹⁴ as a rare type, known as a plerion (also known as a 'filled-centre' or 'Crab-like' SNR). Plerions are young SNRs which have centrally brightened morphologies, and are powered by a central object (known to be a pulsar in the case of the Crab nebula). SGR0526–66 has been associated¹⁵ (controversially) with a SNR (N49) in the Large Magellanic Cloud. Although this remnant is not a plerion, it is relatively young¹⁶ and could contain a pulsar (note that for such a pulsar to power the plerion through rotational energy loss, one would have to reconcile its period with the 8 s periodicity detected in the 5 March 1979 event from this source). Given the apparent radio and optical associations found for the other two SGRs, radio searches of the error region for SGR1900+14 should be an immediate priority. We found three SNRs in current SNR catalogues¹⁷ that are overlapped¹⁸ by this error box: G43.9+1.6, G42.8+0.6 and G39.7–2.0 (=W50). The first two are faint, poorly defined shell remnants, and the latter contains the unusual object SS433. But SS433 is $\sim 0.8^\circ$ (that is, many σ) away from the updated (E. Mazets, personal communication) narrow KONUS error box. Two other SNRs, G40.5–0.5 and G41.1–0.3 are also close to the error box¹⁸. Although the error box is in a complicated region straddling the galactic plane, further studies of this region, and G43.9+1.6 and G42.8+0.6 in particular, are warranted.

Accepting G10.0–0.3 as the site of the SGR1806–20 source, we calculate from the BATSE events (taking a distance of 17 kpc)⁷ that the corresponding burst peak luminosities range between 7×10^{40} and 3×10^{41} erg s⁻¹. This suggests that (even if the source is as close as the Galactic Centre) we now have proof that neutron stars can emit very super-Eddington bursts in low-energy γ -rays, if they are unbeamed; such super-Eddington luminosities are also pointed out in earlier studies¹.

From the fact that reactivation of two of the three known

SGRs has been observed with BATSE, but that no new SGRs have been detected, it follows that the total number, N_{SGR} , of SGR sources in the Galaxy with properties similar to those already known, (including repetition timescales), is probably very limited. A precise limit is difficult to derive in view of gaps in monitoring as well as the differences in sensitivity, sky exposure and trigger criteria of the various experiments capable of observing SGRs. We have estimated a rough upper limit on N_{SGR} by assuming approximately equal sky exposures of BATSE and the experiments with which the SGR sources were detected (~ 12 and ~ 13 yr, respectively). For a series of assumed values of N_{SGR} we calculated the binominal probability that at least two of the known sources are detected again with BATSE, but no new SGR sources are detected, and from this we infer that $N_{\text{SGR}} < 7$ (95% level of confidence). In view of the above mentioned uncertainties we suggest that it is unlikely that our Galaxy contains at any given time more than a few active SGR sources.

The association of SGR1806–20 with a plerion, and of SGR0526–66 with a young SNR, suggests that the burst events

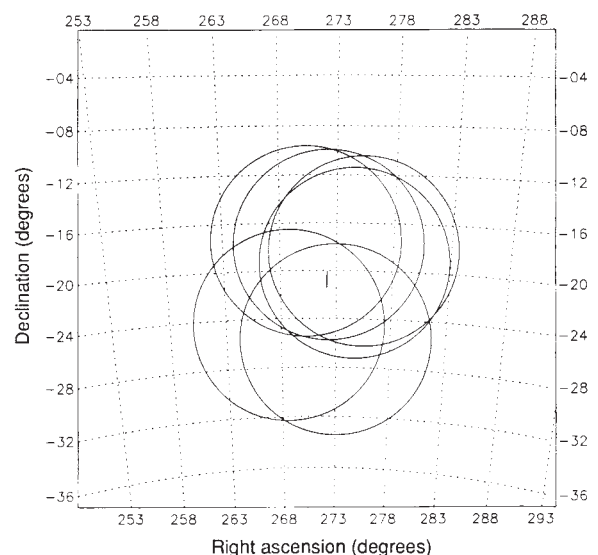


FIG. 2 Sky-map of the region near right ascension 271° , declination -20° . The six large circles represent the locations of the BATSE events. The size of the circles (8°) reflects mainly the statistical errors of the measured positions. Burst locations are obtained from the relative strength of their signals in the relevant subset of the eight identical BATSE Large Area Detectors. The position determination includes the detailed spectral dependence of the angular sensitivity of the detectors, and corrections for photon back-scattering by the Earth's atmosphere. The vertical line represents the 2σ error box ($1.3^\circ \times 4'$) of SGR1806–20 (K. Hurley, personal communication).

originate from young neutron stars with ages of less than $\sim 10^4$ yr. With neutron star birth rates of the order of 1 per 100 yr (refs 19, 20) this would imply that either the active SGR phase lasts less than $\sim 10^3$ yr, or that only a small fraction ($\sim 10\%$) of these young neutron stars become SGR sources. It is currently unknown what special property would turn a young neutron star into an SGR source. Special values of basic parameters such as the neutron star's magnetic field and its rotation rate, or membership of a binary system are obvious possibilities.

If G10.0–0.3 is powered by rotational energy loss of a pulsar it is unlikely that SGRs have a superstrong magnetic field (in excess of 10^{14} G), as suggested before²¹, because this would lead to such rapid spindown that after 5,000 yr (that is, the estimated age of G10.0–0.3)⁷ the remaining spindown power of the neu-

tron star is much too small to power a plerion (this argument is also valid if the neutron star were born with a long spin period). One could, in fact invert the argument, and use the observed⁵ steady (0.5–10 keV) X-ray flux of the plerion to estimate the magnetic field strength of the pulsar believed to reside in G10.0–0.3. Using as an upper limit to the ratio of X-ray luminosity to spindown power the corresponding value observed for the Crab nebula, we find for this field strength an upper limit of $\sim 8 \times 10^{12}$ G, consistent with the magnetic field strengths of young radio pulsars²². Alternatively, the energy source of the plerionic component of G10.0–0.3 may be the magnetic energy of a neutron star; for a field of 10^{14} – 10^{15} G the magnetic energy is sufficient to provide the X-ray luminosity of the plerion for $\sim 10^4$ years. $\square$

Received 12 November 1993; accepted 24 January 1994.

1. Atteia, J.-L. et al. *Astrophys. J.* **320**, L105–L110 (1987).
2. Laros, J. P. et al. *Astrophys. J.* **320**, L111–L115 (1987).
3. Kouveliotou, C. et al. *Astrophys. J.* **322**, L21–L25 (1987).
4. Kouveliotou, C. et al. *Nature* **362**, 728–730 (1993).
5. Tanaka, Y. *IAU Circ. No.* 5580 (1993).
6. Murakami, T. et al. *Nature* **368**, 127–129 (1994).
7. Kulkarni, S. R. & Frail, D. A. *Nature* **365**, 33–35 (1993).
8. Kulkarni, S. R., Frail, D. A., Kassim, N. E., Murakami, T. & Vasisht, G. *Nature* **368**, 129–131 (1994).
9. Norris, J. P., Hertz, P., Wood, K. S. & Kouveliotou, C. *Astrophys. J.* **366**, 240–252 (1991).
10. Norris, J. P. et al. in *Gamma Ray Bursts* (eds Paciesas, W. S. & Fishman, G. J.) 294–298 (AIP 265, Am. Inst. of Phys., New York, 1992).
11. Kouveliotou, C. et al. *Astrophys. J.* **413**, L101–L104 (1993).
12. Laros, J. G. et al. *Nature* **322**, 152–153 (1986).
13. Fenimore, E. E., Laros, J. G. & Ulmer, A. *Astrophys. J.* **418**, 395–397 (1993).

14. Kulkarni, S., Frail, D. A., Murakami, T. & Vasisht, G. *IAU Circ. No.* 5879 (1993).
15. Cline, T. L. *Ann. N.Y. Acad. Sci.* **375**, 314–329 (1981).
16. Vancura, O., Blair, W. P., Long, K. S. & Raymond, J. C. *Astrophys. J.* **394**, 158–173 (1992).
17. Green, D. A. *Publ. astr. Soc. Pacif.* **103**, 209–220 (1991).
18. Reich, W., Fürst, E., Reich, P. & Reif, K. *Astr. Astrophys. Suppl. Ser.* **85**, 633–805 (1990).
19. Strom, R. in *The Lives of Neutron Stars* (Proc. NATO ASI at Kemer, Turkey, September, 1993) (eds Alpar, A., Kiziloglu, U. & van Paradijs, J.) (Kluwer Academic, in the press).
20. Lorimer, D. R., Bailes, M., Dewey, R. J., & Harrison, P. A. *Mon. Not. R. astr. Soc.* **263**, 403–415 (1993).
21. Duncan, R. C. & Thompson, C. in *Gamma Ray Bursters* (Huntsville, October, 1993) (eds Fishman, G. J., Brainerd, J. J. & Hurley, K.) (Am. Inst. of Phys., New York, in the press).
22. Caraveo, P. *Astrophys. J.* (in the press).
23. Boer, M. et al. *Adv. Space Res.* **6**, 97–102 (1986).

ACKNOWLEDGEMENTS. J.V.P. and C.K. thank M. Ruderman for discussions on pulsar magnetic fields.

X-ray identification of the soft γ -ray repeater 1806–20

T. Murakami*, Y. Tanaka*, S. R. Kulkarni†, Y. Ogasaka*, T. Sonobe*, Y. Ogawara*, T. Aoki* & A. Yoshida‡

* Institute of Space and Astronautical Science, 3-1-1 Yoshinodai, Sagami-hara, Kanagawa-ken 229, Japan

† California Institute of Technology, 105–24 Pasadena, California 91125, USA

‡ Institute of Physical and Chemical Research, 2-1 Hirosawa, Wako, Saitama-ken 351, Japan

THE nature of γ -ray bursters—astrophysical sources that emit abrupt bursts of γ -rays—presents a long-standing question in high-energy astronomy. Soft γ -ray repeaters (SGRs) are distinguished from classical γ -ray bursters by the short duration, softer γ -ray spectrum and recurrent activity of their outbursts^{1–6}. Millisecond-scale structure in these bursts suggests that SGRs are compact, and many models invoke neutron stars as the emitting objects^{1,5}. This idea is supported by the association of two SGRs, SGR0526–66 (ref. 6) and SGR1806–20 (ref. 7), with supernova remnants, SNR N49 and the radio nebula G10.0–0.3 respectively. Very recently, Kulkarni et al.^{8,9} have suggested that a compact radio source in G10.0–0.3 corresponds to a young pulsar at the centre of this nebula, and can be identified with SGR1806–20. Here we report the detection of a burst from SGR1806–20 with the X-ray satellite ASCA¹⁰, which allows us to identify the burster with a new X-ray source which we designate AX1805.7–2025. The burst is coincident in time with that detected by the BATSE^{11,12} instrument on board the Compton Gamma Ray Observatory. This result provides strong evidence that SGRs are indeed neutron stars.

On 29 September 1993, the Burst And Transient Source Experiment (BATSE) team discovered that SGR1806–20 became burst-active again¹¹ after almost a decade since the last activity was observed⁴. This motivated us to conduct observations with ASCA for 2 days, from 16:05 UT on 9 October to 14:35 UT on 11 October 1993. The revised location of

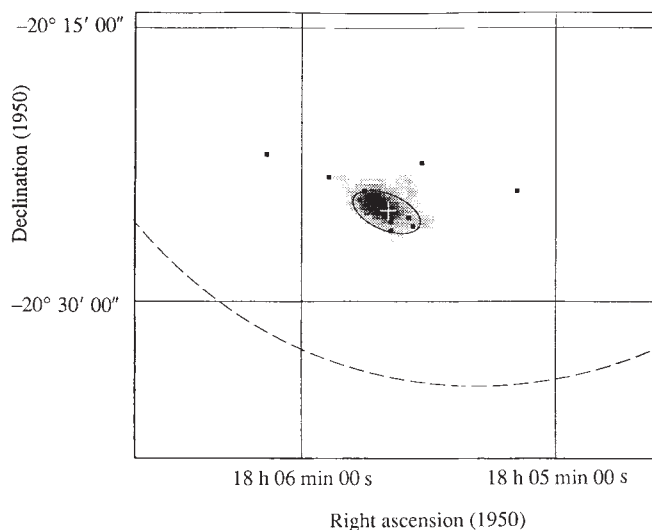


FIG. 1 The GIS image obtained in the position-A observation by the ASCA satellite. A newly discovered X-ray source, designated AX1805.7–2025, appears near the edge of the 50 arcmin-diameter field of view (outer dashed circle) where image distortion is significant. The individual positions of 10 photons from the burst are shown with filled squares. The most probable position of the burst source (white cross) and the 90% confidence error region (ellipse) are indicated. AX1805.7–2025 is well within the error region.

SGR1806–20 is a narrow diamond-shaped region ~ 4 arcmin wide and $1^\circ 20'$ long (K. Hurley, personal communication; see Fig. 1 of ref. 7). The observations were performed at four consecutive positions (positions A to D from north to south) so as to cover the entire error region with the four focal plane detectors; two CCD (charge-coupled device) cameras (solid-state imaging spectrometer, SIS) with a 20×20 arcmin square field of view and two gas scintillation imaging spectrometers (GIS) with a 50 arcmin-diameter circular field of view. The effective integration time was ~ 15 ks each for positions A, C and D, and ~ 30 ks for position B.

Before our analysis, the BATSE team communicated to us (C. Kouveliotou, personal communication) that they had detected a burst of ~ 28 ms duration from SGR1806–20 on 9 October, 22 h 24 min 58 s UT. This was the fifth burst that BATSE detected since the beginning of this activity season. ASCA was pointed to position A, and we were extremely fortunate to have detected this burst¹⁰.

The GIS monitor scalars showed a sharp increase of ~ 170 counts above a steady level of ~ 20 counts per bin for the sum of two GIS rates in a 0.5-s bin which included the time of the burst. However, no significant increase was seen in the SIS rates. Hence, the source of the burst was inside the field of view of GIS but outside the much smaller field of view of SIS.

Each GIS event is processed on board, and the position on the focal plane and energy are determined for each qualified event. These data are temporarily stored in a 4-stage memory cache before read out. For the observation mode employed during this pointing (medium bit-rate mode), read out was performed every 31.25 ms. Because of a short duration of the burst (~ 28 ms) and the limited memory capacity, the position and energy data of only 10 photons (5 for each GIS) out of the total ~ 170 incident photons were recorded.

In the same observation, a previously uncatalogued source was discovered which is clearly visible in the GIS image shown in Fig. 1. The individual positions of 10 burst photons are plotted in the same GIS image. The image suffers a considerable distortion (oval-shaped) near the rim of the focal plane. This is a consequence of the off-axis aberration in the ASCA telescope optics and the image nonlinearity of GIS for objects close to the edge of the field of view. These effects have been incorporated in determining the burst location by employing a measured off-axis point response function (from an observation of the Crab Nebula at about the same off-axis angle). We have verified that the distribution of the 10 burst photons is consistent with a point source. The most probable position of the burst source on the focal plane was determined by finding the best fit to the point response function. The 90% confidence error region, as derived from our Monte Carlo simulations, is $\sim 2 \times 4$ arcmin (Fig. 1), and coincides with the newly discovered X-ray source AX1805.7–2025 (see below).

This source was observed again in a subsequent pointing (position B) within the field of view of both SIS and GIS, which allowed us a more precise determination of its location. An SIS image is shown in Fig. 2. The celestial coordinates of the source are right ascension (RA) 18 h 05 m 41 s, declination (dec.) $-20^\circ 25' 07''$ (equinox 1950.0) with a systematic error circle radius of 1 arcmin. We designate the source AX1805.7–2025. The X-ray spectrum can be expressed by a power law of a photon index 2.2 ± 0.2 (90% confidence limit) with a heavy absorption by a column of $\sim 6 \times 10^{22}$ H atoms cm^{-2} . The flux in the range 2–10 keV is $\sim 6 \times 10^{-12}$ erg cm^{-2} s^{-1} as observed, and 1×10^{-11} erg cm^{-2} s^{-1} or 0.4 mCrab after correction for the absorption. The intensity remained constant over two days, and no change was observed before and immediately after the burst.

The region containing AX1805.7–2025 had been previously surveyed, and the source was undetected. The upper limits are 4.9×10^{-11} erg cm^{-2} s^{-1} (September 1977) and 2.2×10^{-11} erg cm^{-2} s^{-1} (March 1978) from the HEAO-1 satellite observations¹³, and 1×10^{-11} erg cm^{-2} s^{-1} (September 1988) from the Ginga satellite observation¹⁴. All these upper limits are above the ASCA result, yet one can say from these results that no evidence exists for the source to become significantly brighter than the present intensity level.

Recently, Kulkarni *et al.* conducted a multi-band radio obser-

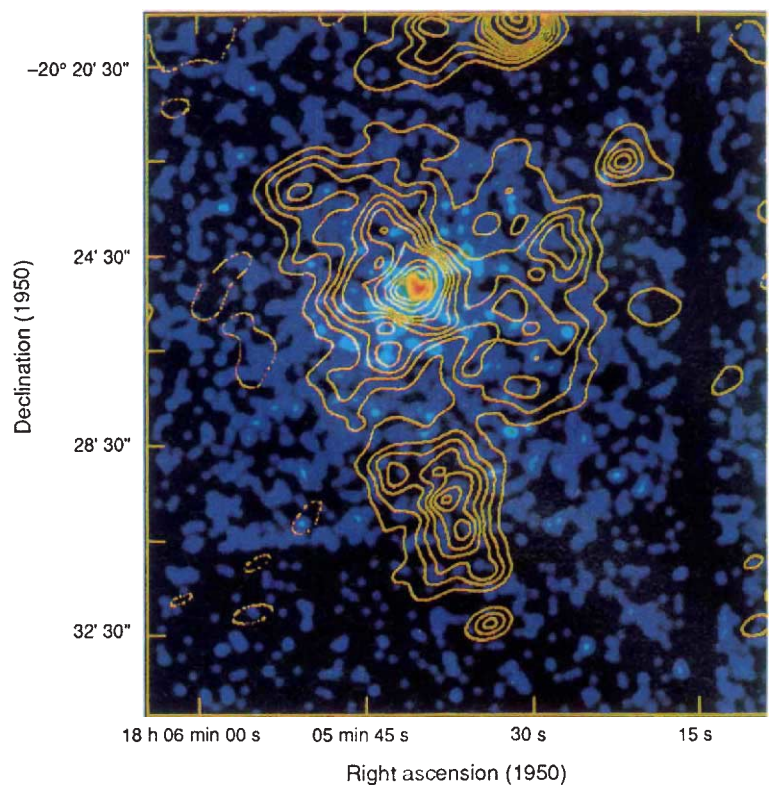


FIG. 2 The SIS X-ray image of AX1805.7–2025 obtained in the position-B observation by ASCA which is superimposed on the radio contours obtained by the VLA observation⁹. The faint 'X'-shaped feature is an artefact due to the response of the X-ray telescope. The source location coincides with the centroid of the compact radio nebula in SNR G10.0–0.3 (refs 8, 9). Courtesy D. A. Frail at VLA Socorro.

vation of SNR G10.0–0.3 at the Very Large Array (VLA)^{8,9}. In the images obtained on 14 September 1993, they found a compact nebula with the centroid at RA 18 h 05 min 41.76 s, dec. $-20^\circ 25' 13''$ (equinox 1950.0), superimposed on an extended plateau of emission. They argue that the morphology indicates a 'plerion', a centrally brightened pulsar-powered synchrotron nebula. Remarkably, AX1805.7–2025 is coincident with the centroid of the compact radio nebula which is shown in Fig. 2.

We may conclude that AX1805.7–2025 is the burst source, not only from the positional coincidence but also from other unique features described below. The X-ray source AX1805.7–2025 is associated with a radio plerion which is rare by itself. The radio, as well as X-ray, spectra are consistent with a synchrotron source such as the Crab Nebula. In addition, the ratio of the observed radio flux⁹ to the X-ray flux is found to be only a factor of 2–3 greater than that for the Crab Nebula, despite a large frequency separation (nine orders of magnitude) between the radio and X-ray frequency bands. Although the X-ray source appears point-like, X-rays may come from a compact nebula of size not much greater than 1 arcmin, which would not have been resolved with the angular resolution of the ASCA telescopes. The similarity of the radio/X-ray flux ratio is consistent with the idea that the source is powered by a pulsar from the radio to X-ray bands.

Thus, the pulsar model provides the most straightforward interpretation of our data. Kulkarni *et al.*⁹ have pointed out that G10.0–0.3 is very unusual for a plerion, but bears a striking resemblance to the radio nebula around Cir X-1 (ref. 15). If we assume a distance of 15 kpc (that used by Kulkarni *et al.*⁹) the X-ray luminosity of the persistent source is $\sim 3 \times 10^{35}$ erg cm^{-2} s^{-1} in the 2–10 keV band. Such a low luminosity, if steady at this level, would be incompatible with systems like Cir X-1. We therefore consider the pulsar model to be most viable, albeit one

in an unconventional plerion, in which both burst-like and steady emission from the pulsar power the surrounding nebula. Clearly, the most decisive evidence for a pulsar is the detection of pulsations. We are in the process of searching for pulsation in the data. □

Received 17 November 1993; accepted 11 January 1994.

1. Liang, E. P. & Petrosian, V. (eds) *Gamma-Ray Burst* (AIP Conf. Proc. No. 141, Am. Inst. of Phys., New York, 1986).
2. Kouveliotou, C. et al. *Nature* **362**, 728–730 (1993).
3. Atteia, J.-L. et al. *Astrophys. J.* **320**, L105–L110 (1987).
4. Laros, J. P. et al. *Astrophys. J.* **320**, L111–L115 (1987).
5. Kouveliotou, C. et al. *Astrophys. J.* **322**, L21–L25 (1987).
6. Cline, T. L. et al. *Astrophys. J.* **255**, L45–L48 (1982).
7. Kulkarni, S. A. & Frail, D. A. *Nature* **365**, 33–35 (1993).
8. Kulkarni, S. A., Frail, D. A., Kassim, N. E. & Murakami, T. *IAU Circ. No. 5879* (1993).
9. Kulkarni, S. A., Frail, D. A., Kassim, N. E., Murakami, T. & Vasishth, G. *Nature* **368**, 129–131 (1994).
10. Tanaka, Y. et al. *IAU Circ. No. 5880* (1993).
11. Kouveliotou, C. et al. *IAU Circ. No. 5875* (1993).
12. Kouveliotou, C. et al. *Nature* **368**, 125–127 (1994).
13. Norris, J. P., Hertz, P., Wood, K. S. & Kouveliotou, C. *Astrophys. J.* **366**, 240–252 (1991).
14. Murakami, T., Nishimura, J., Kawai, N., Cooke, B. A. & Kato, M. *Astr. Astrophys.* **227**, 451–455 (1990).
15. Stewart, R. T., Caswell, J. L., Haynes, R. F. & Nelson, G. J. *Mon. Not. R. astr. Soc.* **261**, 593–598 (1993).

ACKNOWLEDGEMENTS. We thank C. Kouveliotou and the BATSE team for sending a quick notice and providing prompt information of bursts from SGR1806–20, and the ASCA team members for support. We also thank D. Helfand, H. Inoue, F. Nagase and R. Mushotzky for discussions. S.R.K. is supported by NASA, the US NSF and the Packard Foundation. This work was partially supported by a Grant-in-Aid from the Japanese Ministry of Education, Science and Culture.

The radio nebula of the soft γ -ray repeater 1806–20

S. R. Kulkarni*, D. A. Frail†, N. E. Kassim‡, T. Murakami§ & G. Vasishth*

* Division of Physics, Mathematics and Astronomy 105-24, California Institute of Technology, Pasadena, California 91125, USA
† National Radio Astronomy Observatory, PO Box 0, Socorro, New Mexico 87801, USA

‡ Remote Sensing Division, Naval Research Laboratory Code 7215, Washington DC 20375-5351, USA

§ Institute for Space and Astronautical Science, 3-1-1 Yoshinodai, Sagami-hara, 229 Kanagawa, Japan

AN important clue to the nature of soft γ -ray repeaters (SGRs), which emit recurrent bursts of γ -rays, has been provided by the association of two of the three known SGRs with supernova remnants^{1,2} (SNRs). Here we present radio images of the non-thermal radio nebula G10.0–0.3, a supernova remnant which has been associated previously² with SGR1806–20 (refs 3, 4). Our images show that the nebula is a plerion (that is, the radio emission is synchrotron radiation powered by a central pulsar), as revealed by the observation of a hierarchy of nested shells and a bright central peak. The recent detection⁵ of an X-ray point source coincident with the radio peak and of a hard X-ray burst^{5,6} from G10.0–0.3 confirms the SGR–SNR association. We propose that SGR1806–20 is an isolated pulsar that emits both steady and impulsive winds of relativistic particles, which together power the nebula. We suggest that the offset from the centres of the SNRs observed for both this object and SGR0526–66 (ref. 1), requiring high velocities of the pulsars, provides a clue to their formation mechanism.

The new Very Large Array (VLA) observations were obtained on 14 September 1993 with the VLA in the “DnC” configuration. Our previous observations² showed that G10.0–3 was an amorphous non-thermal nebula, consistent with an old SNR, but these observations left open questions about the morphology of the remnant and whether the SGR was in any way responsible for the non-thermal nebular emission. Multifrequency observations were conducted at 0.3, 1.4, 4.8 and 8.4 GHz to gain a better understanding of the nebula by the usual diagnostics of

continuum radio astronomy: morphology and spectral index α (defined from the dependence of flux density S on frequency ν : $S(\nu) \propto \nu^\alpha$).

At the lowest frequencies, a nebula of size 9×6 arcmin² is apparent (Fig. 1). The total flux in these essentially fully sampled images is 3.3 Jy and 1.3 Jy at 0.3 GHz and 1.4 GHz, respectively, giving a spectral index $\alpha = -0.6$. At 1.4 GHz a plateau of emission with a full radius of 40 arcsec can be seen located slightly northeast of the nominal centre of G10.0–0.3. At the higher resolution of the 4.8 and 8.4 GHz images, this plateau is resolved into a compact core region (with a size of a few arcseconds) and complex filamentary structure. This central plateau region also has $\alpha = -0.6$, as measured between 1.4 GHz and 4.8 GHz.

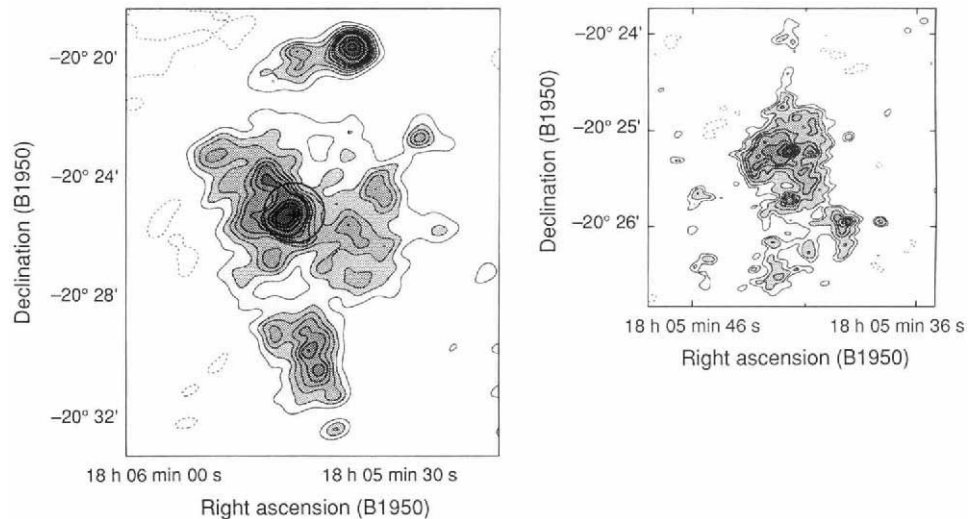
The limb-brightened emission that is characteristic of shell-type SNRs is conspicuously absent in these images. The observed morphology, with strong emission from the central regions, is typical of a plerion⁷. Plerions are SNRs that are believed to be powered by a central source of relativistic particles. The hierarchical structure, culminating in a central peak (Fig. 1), suggests that a compact object presumably corresponding to SGR1806–20 is located at this peak (the ‘radio core’). Our preliminary assertion⁸ that this was so has been confirmed by X-ray observations with the ASCA satellite⁵. The precise VLA position⁸ of the peak has led to other multi-wavelength observations^{9,10}.

As the repository of the energetic emission from the SGR, the nebula serves as a convenient calorimeter. The radio observations reported here, along with the X-ray observations⁵, allow us to constrain the energetics and age of the SGR. We consider two alternative neutron-star models for powering the radio nebula G10.0–0.3: (1) An isolated radio pulsar whose spin-down luminosity is primarily through a relativistic wind. Examples include a synchrotron nebulae powered by young pulsars, such as the Crab Nebula. Less than 10% of the SNRs are classified¹¹ as plerions. (2) An accreting neutron star with radio nebulae such as the ones associated with Cir X-1 (ref. 12). In most accreting neutron-star binaries, the accretion power appears as X-ray emission with no substantial radio emission. However, a small and rare subgroup, including Cir X-1 and SS 433 (ref. 13), power a bright radio nebula. In this group, some fraction of the accretion power is harnessed to accelerate particles to relativistic energies. Strictly speaking, both of these types of nebulae are plerions according to the morphological criterion discussed above. However, here we will reserve the term ‘plerion’ for nebulae powered by (or strongly suspected to be powered by) pulsars. We will refer to nebulae such as those around Cir X-1 or SS 433 as ‘accretion-powered’.

Neither of these two models can fully explain all of the observations. First, G10.0–0.3 has a spectral index of -0.6 , distinctly steeper than the range $\alpha = 0$ to -0.3 characteristic of plerions¹¹. We note however, that older plerions show (unexplained) spectral steepening at centimetre wavelengths^{14,15}, so this objection is not fatal. An accurate spectrum of G10.0–0.3 at metre wavelengths would be most helpful. Second, at a distance of $15d_{15}$ kpc the 9×6 arcmin² nebula would be 39×26 pc², making it the largest known plerion; uncertainties in the distance are parameterized by the dimensionless parameter d_{15} . These dimensions, although large, are similar to two other old, low-surface-brightness plerions, Vela X and G74.9+1.2 (ref. 7).

The radio properties of G10.0–0.3 more closely resemble that of the radio nebula¹² around the X-ray binary Cir X–1. They have similar sizes, spectral indices and morphology. There are, however, at least two objections to the Cir X-1 model. First, the persistent X-ray luminosity, $L_x \approx 3 \times 10^{35} d_{15}^2$ erg, is considerably below the Eddington luminosity. It is our observation that only those binaries with super-Eddington accretion episodes have radio nebula (P. Predehl and S.R.K., manuscript in preparation) for example SS 433, Cir X-1 and perhaps Cyg X-3. Second, these rare binaries also show dramatic radio flaring events that are related to the X-ray bursts. VLA images of the plateau/core

FIG. 1 Radio images of the supernova remnant G10.0-0.3 at 1.4 GHz (left) and 8.4 GHz (right). Both images were made at the Very Large Array (VLA) in the "DnC" configuration, each with ~ 35 min of integration time. The 1.4-GHz image has a resolution of 31 arcsec, whereas the 8.4-GHz image has a resolution of 5.5 arcsec. Contour intervals for the 1.4 GHz image are plotted at $-3, 2, 4, 6, 8, 10, 12, 14, 20, 25$ and 30 times the r.m.s. noise level of $1.56 \text{ mJy beam}^{-1}$, with the negative contour indicated by dashed lines. Contour intervals for the 8.4 GHz image are plotted at $-3, 2, 3, 5, 7, 9, 11, 15, 20, 25$ and 30 times the r.m.s. noise level of $0.057 \text{ mJy beam}^{-1}$. A circle centred on the steady ASCA X-ray source⁵ is shown in the 1.4-GHz image; the radius of 1 arcmin is the estimated uncertainty of the X-ray source. Note that VLA images obtained in a given configuration are sensitive to a restricted range of angular scales, from λ/B_{max} to λ/B_{min} where B_{max} and B_{min} are the maximum and minimum antenna separations and λ is the wavelength of observations; for the observations reported here, $B_{\text{min}} = 35 \text{ m}$ and $B_{\text{max}} = 1,900 \text{ m}$. The low-frequency image has a large field of view and



is good for obtaining the overall view of the source; the high frequency image is suited to the study of compact high-surface-brightness features. With these caveats, the morphology of the new images can be readily interpreted in terms of a compact source which is powering the entire nebula and is also the SGR. The peak of the core is right ascension $18 \text{ h } 05 \text{ min } 41.7 \text{ s}$ and declination $-20^\circ 25' 12''$ (equinox 1950), and this coincides with the steady ASCA X-ray source.

region obtained at 8.4 GHz on 3 October 1993, 4 days after the first γ -ray burst⁶, show no significant ($\lesssim 10\%$) changes in the flux, as compared to the image obtained from the September data. Third, at the radio peak there is no bright infrared source⁹, as is found towards Cir X-1, Cyg X-3 or SS 433.

To summarize, the radio spectral index and the nested morphology of G10.0-0.3 favour a Cir X-1 model, whereas the plerion model is favoured by the X-ray data and the absence of radio variability. We therefore propose a hybrid model which incorporates elements of both. We propose that the persistent luminosity is due to the powering of a plerionic nebula by a middle-aged pulsar like Vela which, for unknown reasons, undergoes super-Eddington bursts of particles. The particles are probably not injected in a spherical wind, but in collimated outflows or jets, as in sources like SS 433 or Cir X-1. Indeed a hint of such a flow is visible in the 8-GHz image (Fig. 1). If our conjecture is correct, then in the past the time-averaged burst particle luminosity must have been much greater than the persistent particle luminosity. (At present, the time-averaged burst photon luminosity appears to be less than the persistent photon luminosity). Thus in our hybrid model, the burst mechanism is not an insignificant drain on the energy budget of the pulsar.

Applying the usual synchrotron emission formulae¹⁶ and assuming low- and high-frequency cutoffs of 10 MHz and 100 GHz, we estimate a minimum energy in particles and magnetic field to be $E_0 \approx 7 \times 10^{49} d_{15}^{1/7} \text{ erg}$. This minimum energy is slightly higher than that for most¹⁷ plerions but not unusually so. Equating E_0 to the initial rotational energy of the pulsar, $2\pi^2 I P_1^{-2}$, we infer the initial period $P_1 \lesssim 16 \text{ ms}$; here I is the moment of inertia. Given the ASCA Crab-like spectrum (that is, $\alpha \approx -1$) of the X-ray source and the radio spectral index ($\alpha = -0.6$), we conclude that, as in the Crab nebula, the electromagnetic spectrum of G10.0-0.3 has a synchrotron spectral break somewhere between the radio and the X-ray bands.

The ASCA hard X-ray source is the compact nebula formed by the shock of the relativistic wind. We predict a fainter, soft X-ray extended nebula. As with other plerions, we assume¹⁸ a rough inefficiency of $\eta \approx 10^{-2} \eta_{-2}$ between the X-ray emission and $\dot{E}$ the spin-down luminosity of the pulsar; here η_{-2} is a dimensionless quantity that allows us to parametrize the uncertainty in η . Thus $\dot{E} \approx 3 \times 10^{37} \eta_{-2}^{-1} d_{15}^2 \text{ erg s}^{-1}$, leading to a characteristic nebular age, $t_n \approx E_0/\dot{E} \approx 8 \times 10^4 \eta_{-2} d_{15}^{3/7} \text{ yr}$. Because

adiabatic losses could have been considerable, we note that t_n is an underestimate.

The 2-arcmin offset location of the pulsar with respect to the centroid of G10.0-0.3 provides a lower limit to t_p , the age of the pulsar. Assuming that the pulsar has a velocity of $v < 10^3 \text{ km s}^{-1}$, the highest measured pulsar velocity¹⁹, we obtain $t_p \geq 6 \times 10^3 \text{ yr}$. t_p cannot be much larger than this value because it is well known that plerions fade quite rapidly²⁰, $t_p \lesssim 10^4 \text{ yr}$. Thus the pulsar has a large velocity, $v \gtrsim 500 \text{ km s}^{-1}$. As $t_n \gg t_p$, we conclude significant $\dot{E}$ evolution has taken place. Assuming a simple magnetic dipole spin-down model, we constrain the magnetic field strength, $B \lesssim 10^{13} \text{ G}$; here we have made the reasonable assumption that the current period is significantly larger than P_1 .

The properties of the hypothesized pulsar are not different from those of a middle-aged pulsar like the Vela pulsar. SGRs are, however, very rare²¹. Thus we are forced to the conclusion that only a very few, unusual pulsars evolve to SGRs. In this context, it should be noted that both SGR1806-20 and SGR0526-66 (associated¹ with SNR N49) are offset from the centres of their SNRs. Thus both neutron stars have substantial space motion, $v \gtrsim 500 \text{ km s}^{-1}$, suggesting that a connection may exist between the SGR phenomenon and high velocity. We speculate that these two are manifestations of a more fundamental process. For example, models²² of SGRs that advocate peculiar magnetic field configurations or extraordinary field strengths ($B \approx 10^{15} \text{ G}$) can conceivably generate large peculiar velocities, either at birth or shortly thereafter. Massive internal changes could be effected when the field tries to reach a more regular distribution or reconnects; these changes might then give rise to the soft γ -ray bursts. In any case, a successful model for SGRs must explain why SGRs have high velocities. $\square$

Received 8 November 1993; accepted 25 January 1994.

1. Cline, T. L. et al. *Astrophys. J.* **255**, L45-L48 (1982).
2. Kulkarni, S. R. & Frail, D. A. *Nature* **365**, 33-35 (1993).
3. Atteia, J.-L. et al. *Astrophys. J.* **320**, L105-L110 (1987).
4. Laros, J. G. et al. *Astrophys. J.* **320**, L111-L115 (1987).
5. Murakami, T. et al. *Nature* **368**, 127-129 (1994).
6. Kouveliotou, C. et al. *Nature* **368**, 125-127 (1994).
7. Weiler, K. W. & Sramek, R. A. *Rev. Astr. Astrophys.* **26**, 295-341 (1988).
8. Kulkarni, S. R., Frail, D. A., Kassim, N. E., Murakami, T. & Vasisht, G. *I.A.U. Circ. No. 5879* (1993).
9. Kulkarni, S. et al. *I.A.U. Circ. No. 5883* (1993).

10. Cooke, B. A. *Nature* **366**, 413–414 (1993).
11. Green, D. A. *Publ. astr. Soc. Pacif.* **103**, 209–220 (1991).
12. Stewart, R. T., Caswell, J. L., Haynes, R. F. & Nelson, G. J. *Mon. Not. R. astr. Soc.* **261**, 593–598 (1993).
13. Margon, B. *Rev. Astr. Astrophys.* **22**, 507–536 (1984).
14. Morsi, H. W. & Reich, W. *Astr. Astrophys. Suppl. Ser.* **71**, 189–195 (1987).
15. Green, D. A. & Scheuer, P. A. G., *Mon. Not. R. astr. Soc.* **258**, 833–840 (1992).
16. Pacholczyk, A. G. *Radio Astrophysics*, Ch. 7, Freeman, San Francisco, 1970.
17. Helfand, D. J. & Becker, R. H. *Astrophys. J.* **314**, 203–214 (1987).
18. Seward, F. D. & Wang, Z. R. *Astrophys. J.* **332**, 199–205 (1988).
19. Harrison, P. A., Lyne, A. G. & Anderson, B. *Mon. Not. R. astr. Soc.* **261**, 113–124 (1993).
20. Braun, R., Goss, W. M. & Lyne, A. G. *Astrophys. J.* **340**, 355–361 (1989).
21. Kouveliotou, C. *Astrophys. J.* **392**, 179–186 (1992).
22. Duncan, R. C. & Thompson, C. *Astrophys. J.* **392**, L9–L13 (1992).

ACKNOWLEDGEMENTS. We thank M. Rupen for assistance with the October observations at the VLA. S.R.K. gratefully acknowledges discussions with R. Blandford. D.A.F. thanks R. Hjellming for discussions. S.R.K.'s work is supported by grants from NASA, US NSF and the Packard Foundation. The VLA is operated by Associated Universities Inc. under the cooperative agreement with the US National Science Foundation. Basic research in Radio Interferometry at the Naval Research Laboratory is supported by the US Office of Naval Research.

Direct identification of atomic and molecular adsorption sites using photoelectron diffraction

Ph. Hofmann, K.-M. Schindler, S. Bao,
A. M. Bradshaw & D. P. Woodruff*

Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4–6,
14195 Berlin Dahlem, Germany

* Department of Physics, University of Warwick, Coventry CV4 7AL, UK

THE determination of the adsorption geometry of molecules and molecular fragments on single-crystal metal surfaces is central to our understanding of heterogeneous catalysis. Adsorbate structures can be determined precisely by techniques such as low-energy electron diffraction and photoelectron diffraction. Such methods suffer, however, from the rather inefficient approach to data analysis: the diffracted intensities are compared with simulated data for a trial structure, which is successively modified by trial and error until good agreement is achieved. Following a suggestion by Barton that a photoelectron angular distribution may be regarded as a photoelectron hologram^{1,2}, attention has been focused recently on simpler methods for adsorbate structure determination which give real-space information directly^{3–7}. We demonstrate here a direct method for determining both the adsorption site and the adsorbate–substrate separation from photoelectron diffraction data⁸. We illustrate the method using two adsorption systems: CO on Cu(110) and OCH₃ on Cu(111). Our results are consistent with those determined for the same systems by a related method⁷, which requires a considerably larger data set and provides bond lengths at lower precision. For adsorbates on simple surfaces, we therefore propose that our approach provides an accurate and easily implemented method for structure determination.

Although the notion of photoelectron holography (imaging of atomic positions by the reconstruction of single-wavelength angular distributions) is at first sight rather attractive, most holographic reconstructions so far performed^{13,14} suffer from poor resolution, misplaced ‘atoms’ and strong spurious peaks due to effects such as limited data range and the neglect of scattering phase shifts. More promising reconstructions have only been achieved using extremely large simulated data sets determined at several electron kinetic energies^{15,16}. There have only been few reports of the application of holography to adsorption systems², although Barton’s original paper¹ was concerned with this topic.

Two direct methods^{3–7} based on Fourier transforms of wavelength-dependent measurements have recently enjoyed some success in determining adsorbate sites. Moreover, they give emitter/nearest-neighbour separations to within a few tenths of an ångström. Both methods still require, however, relatively large data sets, for example 48 scanned-energy-mode photoelectron

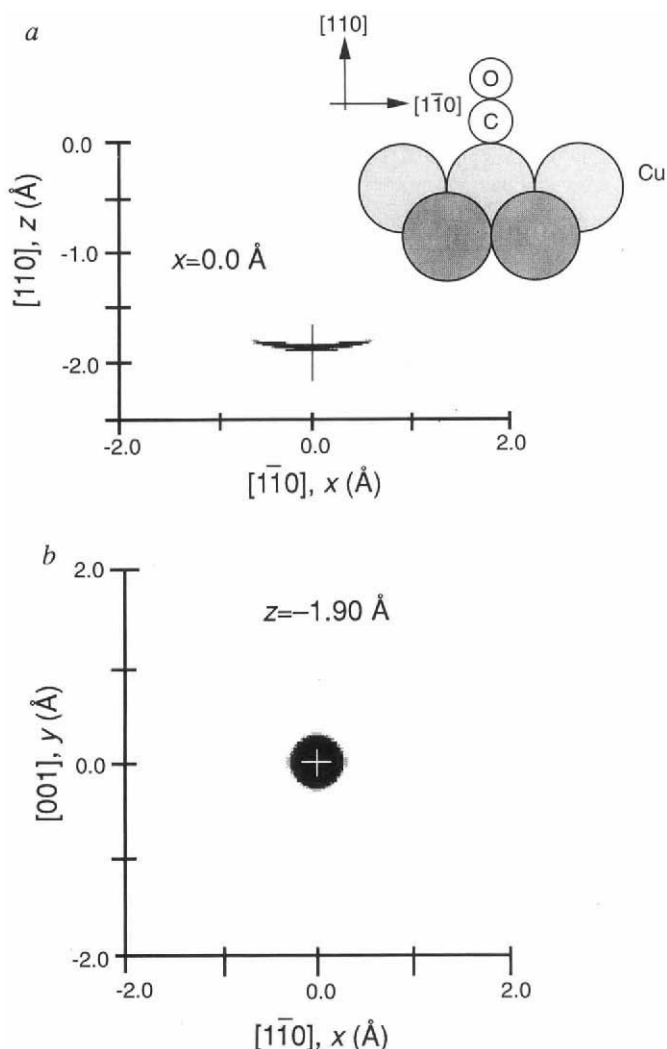


FIG. 1 The coefficient $C(\mathbf{r})$ for the system Cu(111)–CO. *a*, The zy plane at $x=0.0$ Å, and *b*, the xy plane 1.9 Å below the emitter. The inset shows the resulting structural model. The crosses mark the positions of the atoms derived from multiple-scattering spherical-wave calculations.

diffraction spectra in ref. 4, typically 20–25 in ref. 7. Our new approach to the transformation of scanned-energy-mode data into real-space images of the nearest-neighbour surface atoms has previously been applied only to simulated photoelectron data⁸. These results, as well as the more significant test with real experimental photoelectron diffraction data described here, indicate that the approach represents an important advance over holography and other recent direct methods in terms of both accuracy and the size of the data sets used. This is due to the fact that as well as identifying the 180° backscattering direction for the nearest-neighbour substrate atoms—the salient feature of our previous method^{6,7}—scattering phase shifts are also included.⁹

The basic idea of the inversion is to calculate the projection of an experimentally determined photoelectron diffraction modulation function, $\chi_{\text{ex}}(\mathbf{k})$, on a theoretical modulation function, $\chi_{\text{th}}(\mathbf{k}, \mathbf{r})$, which would be caused by a single scatterer (a substrate atom) at $\mathbf{r}$:

$$c(\mathbf{r}) = \int_{\mathbf{k}} \chi_{\text{ex}}(\mathbf{k}) \chi_{\text{th}}(\mathbf{k}, \mathbf{r}) d\mathbf{k}$$

where c is the projection coefficient and $\mathbf{k}$ is the wavenumber of the outgoing photoelectron wave in the direction of the detector.

Such a projection has a simple geometric interpretation in vector algebra and is used here in a more generalized way for functions. The $c(\mathbf{r})$ are then combined for several experimental modulation functions (taken in different emission directions) to give a coefficient $C(\mathbf{r})$ derived from the whole data set⁶. We may expect strong maxima in $C(\mathbf{r})$ at the actual positions $\mathbf{r}$ of the nearest-neighbour surface atoms and can hence determine the adsorption structure. Here, we test this method on experimental data sets for two adsorption systems.

The first example is provided by the adsorption of CO on Cu(110). Although the structure of this particular system has not been determined previously, it is very likely that the molecule is bonded by the carbon atom to the surface in an atop site (that is, directly above a copper atom in the first layer) on Cu(100) (ref. 10). In our experiments scanned-energy-mode photoelectron diffraction data from the C 1s core level in normal emission as well as in the 20° and 40° off-normal directions in both [001] and [110] azimuths were measured on the HE-TGM 1 monochromator at the synchrotron radiation source BESSY. (An O 1s or C 1s modulation function recorded in one emission direction at 190 photon energies takes ~1 h to collect, so that the data can easily be obtained in a day.) For the purposes of our direct transformation, the data set was extended by a factor of two by making use of the 2-fold rotational symmetry axis perpendicular to the surface. The coefficient $C(\mathbf{r})$ was calculated in a cartesian co-ordinate system with the z-axis perpendicular to the surface and the carbon atom at the origin. Figure 1a shows $C(\mathbf{r})$ in a plane perpendicular to the surface. Dark areas correspond to high values of $C(\mathbf{r})$. A single structure can be observed ~1.9 Å below the emitting carbon atom. The elongation of the surface atom 'images' is related to the parabolae of constant scattering path length difference³. Figure 1b shows the same structure in a plane parallel to the surface 1.9 Å below the emitter. The crosses indicate the positions of the Cu atoms as obtained from comprehensive multiple scattering calculations with the trial-and-error method. Note that the distance to the nearest-neighbour scatterer determined by this method is very close (within 0.1 Å) to that of the final result. The form of the 'image' does not affect the accuracy because the elongation is tangential to the direction of the nearest neighbour, or backscattering direction. We may therefore conclude that the CO molecule is adsorbed in an atop site on this copper surface with an approximate Cu-C bond length of 1.9 Å (see inset Fig. 1).

The second example is the methoxy species (OCH₃), which adsorbs with its O-C axis perpendicular to the Cu(111) surface^{11,12}. The adsorption site is, however, not known. We have taken O 1s diffraction data for normal emission as well as at polar angles of 30° and 50° in the [110], [121] and [211] azimuths. Again, use has been made of the rotational symmetry in order to extend the data set before performing the direct image reconstruction. The result of the transformation is given in Fig. 2. In a plane perpendicular to the surface in a <211> azimuth (Fig. 2a) a single feature ~1.33 Å below the emitter is found. If we look at $C(\mathbf{r})$ in the plane parallel to the surface for $z = -1.33$ Å, three strong maxima are found which correspond to the Cu nearest neighbours and which therefore indicate that the methoxy species is adsorbed in a 3-fold hollow site (inset). Moreover, the orientation of the 'triangle' of nearest-neighbour surface atoms indicates that the 'face-centred cubic' rather than

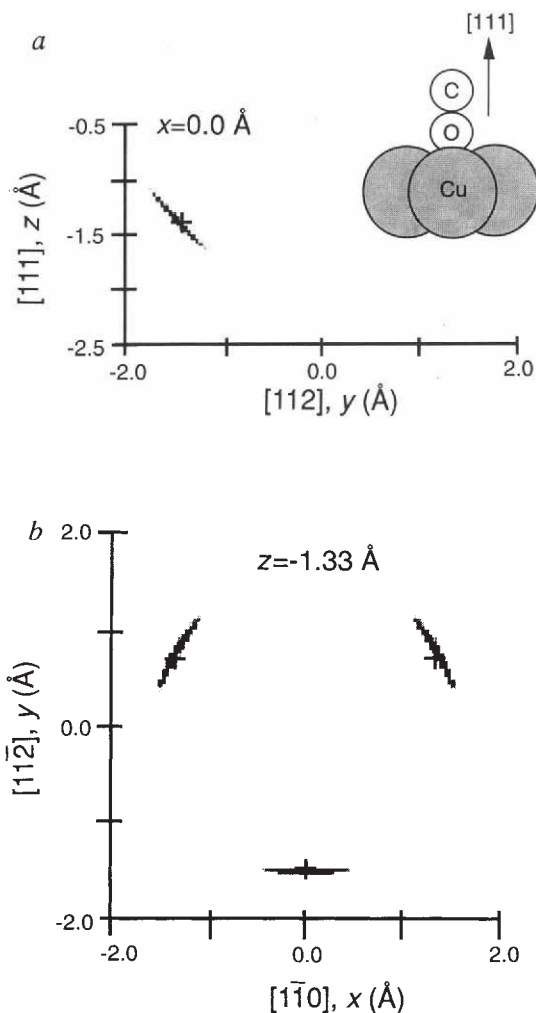


FIG. 2 The coefficient $C(\mathbf{r})$ for the system Cu(111)-OCH₃. a, The xy plane at $x=0.0$ Å, and b, the xy plane 1.4 Å below the emitter. The inset shows the resulting structural model (H atoms not shown). The crosses mark the positions of the atoms derived from multiple-scattering spherical-wave calculations.

the 'hexagonal close packing' 3-fold hollow site is occupied. As in the case of CO mentioned above, the atomic positions agree very well with the results for the full analysis (crosses in Fig. 2).

These examples illustrate that our new direct transformation technique is a powerful tool for adsorption site identification. It has since been applied successfully to over a dozen adsorption systems, even in situations where adsorbate-induced reconstruction occurs. Only a small set of scanned-energy-mode photoelectron diffraction data is required and the calculation times are short. The resulting structural parameters are very good initial estimates, which can be further refined using full multiple-scattering calculations within the framework of the trial-and-error method. The degree of precision attained has so far not been reached with other direct methods. □

Received 1 June 1993; accepted 11 January 1994.

- Barton, J. J. *Phys. Rev. Lett.* **61**, 1356-1359 (1988).
- Thevuthasan, S. et al. *Phys. Rev. Lett.* **70**, 595-598 (1993).
- Tong, S. Y. & Huang, H. *Phys. Rev. Lett.* **64**, 2452-2458 (1992).
- Tobin, J. G., Waddill, G. D., Li, H. & Tong, S. Y. *Phys. Rev. Lett.* **70**, 4150-4153 (1993).
- Wu, H., Lapeyre, G. J., Huang, H. & Tong, S. Y. *Phys. Rev. Lett.* **71**, 251-254 (1993).
- Fritzsche, V. & Woodruff, D. P. *Phys. Rev. Lett.* **64**, 16128-16134 (1992).
- Schindler, K.-M. et al. *Phys. Rev. Lett.* **71**, 2054-2057 (1993).
- Hofmann, Ph. & Schindler, K.-M. *Phys. Rev. Lett.* **67**, 13941-13943 (1993).
- Hu, P. & King, D. A. *Nature* **360**, 656-658 (1992).
- Andersson, S. & Pendry, J. B. *Phys. Rev. Lett.* **43**, 363-365 (1979).
- Ricken, D. E., Somers, J., Robinson, A. & Bradshaw, A. M. *Faraday Discuss. chem. Soc.* **89**,

- 291-300 (1990).
- de Carvalho, A. V., Asensio, M. C. & Woodruff, D. P. *Surf. Sci.* **273**, 381-384 (1992).
- Harp, G. R., Saldin, D. K. & Tonner, B. P. *Phys. Rev. Lett.* **64**, 9199-9202 (1990).
- Herman, G. S., Thevuthasan, S., Tran, T. T., Kim, Y. J. & Fadley, C. S. *Phys. Rev. Lett.* **68**, 650-653 (1992).
- Tong, S. Y., Li, H. & Huang, H. *Phys. Rev. Lett.* **67**, 3102-3105 (1991).
- Barton, J. J. *Phys. Rev. Lett.* **67**, 3106-3109 (1991).

ACKNOWLEDGEMENTS. This work was supported by the German Federal Ministry for Research and Technology (BMFT), the UK Science and Engineering Research Council (SERC) and the ECU SCIENCE programme.

Luminescent anodized silicon aerocrystal networks prepared by supercritical drying

L. T. Canham*, A. G. Cullis*, C. Pickering*,
O. D. Dosser*, T. I. Cox* & T. P. Lynch†

* DRA Malvern, St Andrews Road, Malvern, Worcester WR14 3PS, UK

† BP International Ltd, Group Research and Engineering Centre,
Chertsey Road, Sunbury-on-Thames TW16 7LL, UK

THE preparation of highly porous materials by sol-gel processing followed by supercritical drying was pioneered by Kistler in 1931¹. These materials, called aerogels and comprising a void fraction of more than 90%, are currently a focus of technological interest². Here we show that supercritical drying can be used to prepare highly porous silicon, a material of particular interest because of its luminescent properties³. Porous silicon is prepared conventionally by electrochemical etching, but using supercritical fluid extraction allows us to obtain porosities that would otherwise lead to collapse of the network during drying. Electron microscopy reveals that our materials have a crystalline, columnar structure, and gravimetric and ellipsometric analysis indicates that the porosity exceeds 95%. Our silicon aerocrystals exhibit strong photoluminescence. We predict that supercritical drying should greatly improve those properties of porous silicon films that are critical for optoelectronic applications.

When a wet porous material is allowed to dry by evaporation, the liquid-vapour interfaces formed can generate enormous capillary tensions. These forces can induce cracking, shrinkage and even complete disintegration of delicate solid skeletons⁴. The basic idea of supercritical drying is the removal of liquid from the pores above its critical point so as to avoid liquid-vapour interfaces. By converting the liquid into a supercritical fluid and subsequently a gas, crossing of the liquid-vapour phase boundary is avoided and contraction of the structure being dried is thereby minimized. Such techniques have found widespread use in the processing of biological⁵ as well as ceramic⁶ materials. They have never before, however, been used in porous Si technology. Bulk crystalline Si can be made highly porous by anodization in HF-based solutions, and surface films have been studied since 1956⁷. Although several workers have commented that layers of too high a porosity and thickness degrade visually on drying⁸⁻¹⁰, a means of circumventing this has not previously been suggested.

The (001) Si layers of ultra-high porosity (UHP) were all produced by electrochemical etching in '10% ethanoic HF' as described in detail elsewhere¹¹. Layer UHP1 was formed in a heavily boron doped P⁺ (0.01 Ω cm) Czochralski-grown Si wafer by anodization at 50 mA cm⁻² for 3 min followed by a 15-min chemical dissolution step in the same electrolyte. Layer UHP2 was formed in a P⁺ (0.01 Ω cm) Si wafer similarly anodized, but with a 30-min chemical dissolution step, and layer UHP3 in a P⁺ (0.03 Ω cm) Si wafer anodized at 50 mA cm⁻² for 5 min with a 5-min chemical dissolution. Gravimetric data suggest that, assuming no layer thinning during etching and no layer shrinkage during drying, the average porosities and thicknesses would be 91% and 4.9 μ m for UHP1, 95% and 4.9 μ m for UHP2 and 97% and 6.5 μ m for UHP3. Layer UHP1 was allowed to dry in ambient air. Much of the 36 cm² anodized area crazed heavily during electrolyte evaporation but ~20% remained intact. Layers UHP2 and UHP3, which would both craze completely during air drying, were transferred whilst still wet to a pure ethanol bath. The anodized wafers were cleaved into suitable segments under ethanol and stored in ethanol for 24 h before supercritical drying (denoted by suffix SD) or normal air drying (AD).

Wafer segments were supercritically dried by one of the

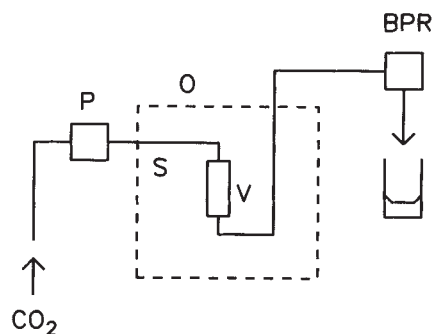


FIG. 1 Flow system and technique used for supercritically drying porous Si films. High-purity CO₂ gas is liquified with a Gilson 303 pump and chilled (−8 °C) Haake F3 pumphead (P). It flows through narrow-bore stainless steel tubing (S) into the high-pressure extraction vessel (V) within a Hewlett-Packard 5880A GC Oven (O). Pressure is controlled with a Jasco back pressure regulator 880-81 (BPR).

METHODS. A segment of UHP2 (see text for nomenclature used to describe these ultra-high porosity Si films) was initially dried slowly in a 50 cm³ Suprex pressure vessel. The ethanol in the pores was replaced by liquid CO₂ by flushing at 2 cm³ min⁻¹ at 18 °C and 1,500 p.s.i.. The level of ethanol remaining was continuously monitored by gas chromatography of the effluent stream. After 3 h of flushing, residual ethanol was at the <10 p.p.m. level. The pressure vessel was then sealed and the temperature ramped up at 1 °C min⁻¹ to 40 °C. The pressure rose gradually to 2,400 p.s.i.. CO₂ in the pores was then supercritical ($T_c = 31.1$ °C, $P_c = 1,073$ p.s.i.) and further flushing was undertaken for 2 h at 1,500 p.s.i. and 40 °C. Finally the CO₂ was vented off over a period of 16 h maintaining a vessel temperature of 40 °C to leave the dried porous Si film. A segment of UHP3 was dried much more rapidly in a smaller (10 cm³) pressure vessel. The ethanol in the pores and vessel was displaced in 30 min, the temperature raised to 40 °C in 15 min, and supercritical CO₂ flushing conducted for only 15 min. Depressurisation at 40 °C was completed in 60 min and the dried sample removed. Additional segments of UHP2 and UHP3 were also dried at an intermediate rate. All supercritically dried films were stored under dry N₂ before characterization.

approaches under development for SiO₂ aerogels^{12,13}. The equipment and techniques used at the BP Supercritical Fluid Laboratories are described in Fig. 1.

Electron microscopy was extensively used to examine the microstructure of dried layers, both scanning electron microscopy (SEM) and high-resolution transmission electron microscopy (HRTEM). Figure 2a–d shows SEM cross-sectional and plan-view images of layer UHP2AD dried in ambient air (Fig. 2a, c) and UHP2SD dried at the lowest rate using supercritical CO₂ (Fig. 2b–d). Whereas the conventionally dried layer is heavily crazed and in some areas has become detached completely from the substrate, no cracks were evident in the supercritically dried segment, even at much higher levels of magnification. A similar dramatic improvement was observed for UHP2SD and UHP3SD supercritically dried at the higher rate, when compared with air-dried UHP2AD and UHP3AD respectively. Figure 2e–f shows the structure of a supercritically dried UHP Si layer, as revealed by HRTEM. Electron-transparent samples were prepared by direct cleavage as described previously¹⁴. Interconnected Si columns of 30–50 Å width are clearly evident due to the open morphology of the structure. The (001) electron diffraction pattern (Fig. 2f) demonstrates the excellent crystallinity of the structure. The strong (220) type spots and weak (400) type spots exhibit limited arcing, indicative of some angular distortions within the crystalline skeleton. But in contrast to luminescent layers dried conventionally¹⁴, neither (311) type nor (111) type rings are discernible. This indicates that the concentration of randomly aligned crystallites resulting from skeleton breakage has been dramatically reduced, despite the structure having even higher porosity than that studied previously¹⁴.

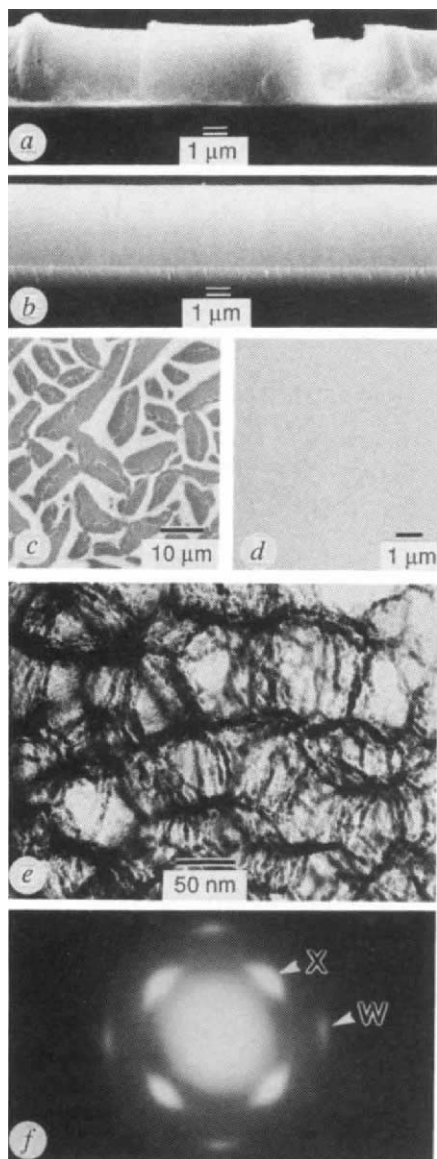


FIG. 2 Electron microscopy data on dried porous and dried ultra-high porosity Si structures. Scanning electron images, taken on a JEOL 6400F microscope, of a porous Si film (UHP2) dried in ambient air (a and c), and when dried with supercritical CO_2 (b and d). Transmission electron plan-view image (e) and diffraction pattern (f) from a (001) oriented flake of UHP2SD taken on a JEM 400EX microscope using 400 keV electrons. In f, spots X are (220) type and W are (400) type.

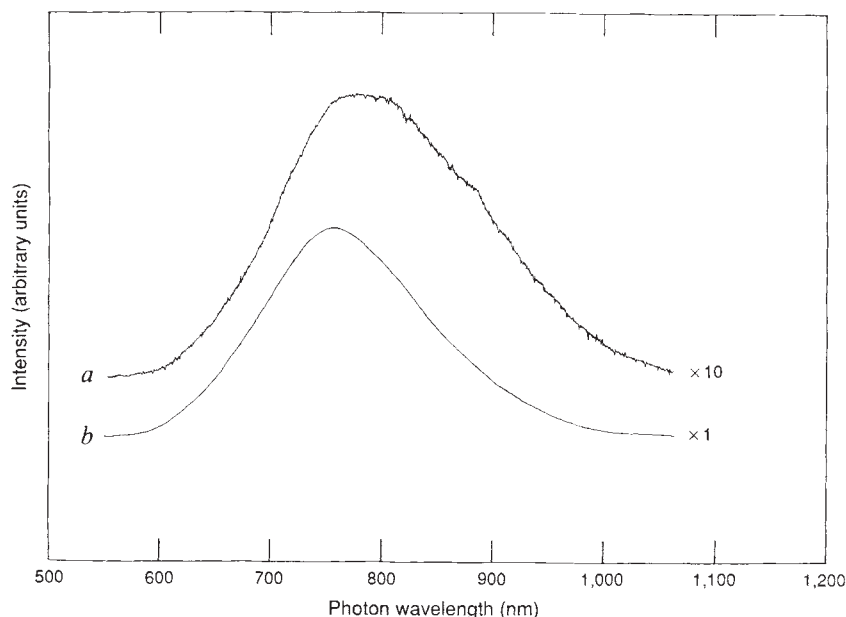


FIG. 4 Room temperature (300 K) photoluminescence spectra from two red-emitting layers of ultra-high porosity silicon that have good structural integrity, a, UHP1AD and b, UHP3SD. Light emission was for Ar laser excitation (0.1 W cm^{-2} at 458 nm), dispersed with a SPEX monochromator and detected with a Ge photodiode (North Coast Inc.). Spectra are not corrected for apparatus response.

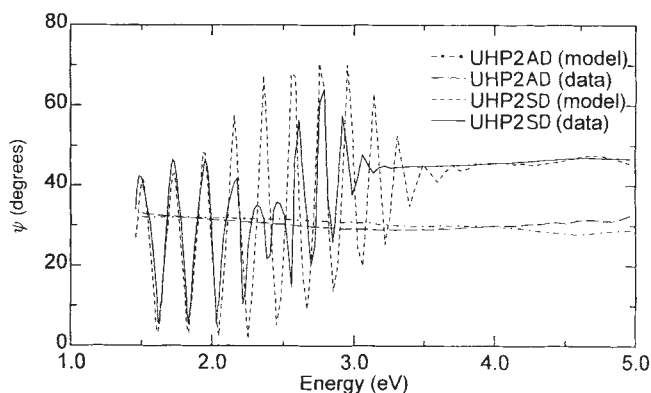


FIG. 3 Ellipsometric parameter, ψ (see text), determined at an angle of incidence of 75° using a SOPRA E546 rotating-polarizer instrument, shown as a function of energy for layers UHP2AD and UHP2SD. Models used are: UHP2AD, Si + 83% voids ($4.7 \mu\text{m}$, no substrate); UHP2SD, Si + 97–92% voids with gradient from surface to interface ($4.98 \mu\text{m}$), Si substrate. The overestimate of the fringe amplitude by the model may be due to the presence of the gradient and/or non-ideal interfaces.

Spectroscopic ellipsometry was used to verify that extremely high porosities had been achieved. Figure 3 compares raw data, in terms of ψ , the amplitude ratio of the reflective coefficients for p- and s- polarized light, for layers UHP2AD and UHP2SD. Marked differences have clearly arisen from the differing drying processes. The data have been fitted using multilayer effective medium models described previously¹⁵. Data for UHP1 (not shown) indicate 91% bulk porosity in agreement with the gravimetric value. The fit to the supercritically dried UHP2SD layer, shown in Fig. 3 indicates a positive porosity gradient ranging from 97% at its surface to 92% in the bulk, in agreement with the higher porosity determined gravimetrically. Thus, the top $0.25 \mu\text{m}$ of this layer is estimated to have an average density of 0.07 g cm^{-3} and a refractive index below 1.10. The data for the conventionally dried layer show no interference fringes, indicating that coherency with the substrate has been lost. These data can be fitted with an average porosity of 83% across a negative gradient, assuming a free-standing layer, which is consistent with cross-sectional SEM results showing lift-off of a large fraction of the layer. This higher density and negative gradient indicate that significant collapse of the porous air-dried structure has occurred.

Figure 4 shows typical photoluminescence spectra (taken at 300 K) for P^+ layers that possess both high porosity and structural integrity. Considerably higher radiative efficiencies, while maintaining film integrity, are achievable with supercritically dried structures (spectrum b) compared with air-dried structures (spectrum a) because smaller skeletons are able to survive the drying process. We have thus demonstrated that supercritical

drying can yield ultra-high porosity luminescent structures that would not otherwise be achievable. In particular, we conclude by stressing that such processing will in general yield porous Si with minimal damage and a much higher degree of perfection than previously attained. This will thereby lead to dramatic improvements in optoelectronic properties. □

Received 20 December 1993; accepted 21 January 1994.

1. Kistler, S. S. *Nature* **127**, 741 (1931).
2. Fricke, J. *Scient. Am.* **256**, 68–73 (1988).
3. Canham, L. T. *Appl. Phys. Lett.* **57**, 1046–1048 (1990).
4. Brinker, C. J. & Scherer, G. W. *Sol Gel Science* Ch. 8 (Academic, San Diego, 1990).
5. Schneider, G. M., Stahl, E. & Wilcke, G. (eds) *Extraction with Supercritical Gases* (Verlag Chemie, Weinheim, 1980).
6. Matson, D. W. & Smith, R. D. *J. Am. Ceram. Soc.* **72**, 871–881 (1989).
7. Uhler, A. *Bell Syst. tech. J.* **35**, 333–347 (1956).
8. Beale, M. I. J., Benjannin, J. D., Uren, M. J., Chew, N. G. & Cullis, A. G. *J. Cryst. Growth* **73**, 622–636 (1985).
9. Friedersdorf, L. E., Searson, P. C., Prokes, S. M., Glembocki O. J. & Macauley, J. M. *Appl. Phys. Lett.* **60**, 2285–2287 (1992).
10. Lehman, V. *Adv. Mater.* **4**, 762–764 (1992).
11. Canham, L. T., Houlton, M. R., Leong, W. Y., Pickering, C. & Keen, J. M. *J. appl. Phys.* **70**, 422–431 (1991).
12. Tewari, P. H., Hunt, A. J. & Lofftus, K. D. *Mater. Lett.* **3**, 363–367 (1985).
13. Rangarajan, B. & Lira, C. T. *J. Supercritical Fluids* **4**, 1–6 (1991).
14. Cullis, A. G. & Canham, L. T. *Nature* **353**, 335–337 (1991).
15. Pickering C., Canham, L. T. & Brumhead, D. *Appl. Surf. Sci.* **63**, 22–26 (1993).

ACKNOWLEDGEMENTS. We thank P. W. Smith for experimental assistance, C. J. Canham for stimulating discussions and M. Atkins for his interest in the work.

Correlations between SO₂ flux and long-period seismicity at Galeras volcano

Tobias P. Fischer*, **Meghan M. Morrissey***,
Marta Lucía Calvache V.*†, **Diego Gómez M.†**,
Roberto Torres C.†, **John Stix‡**
& **Stanley N. Williams***

* Department of Geology, Arizona State University, Tempe, Arizona 85287–1404, USA

† Observatorio Volcanológico del Sur, Ingeominas, Apartado Aéreo 1795, Pasto, Colombia

‡ Département de Géologie, Université de Montréal, Montréal, Québec H3C 3J7, Canada

THE 14 January 1993 eruption of Galeras volcano, in Colombia, which killed six scientists and three tourists¹, was followed by a larger eruption on 23 March 1993². Both eruptions were preceded by episodes of long-period seismicity. The source of long-period seismic events has been modelled extensively^{3–8}, as the resonance within a fluid-filled crack induced by pressure transients in the fluid^{5–7}. Here we use the SO₂ flux from Galeras volcano, measured remotely, to infer the degassing history during the episode of long-period events preceding the 23 March eruption. SO₂ flux and long-period seismicity have been monitored separately elsewhere to forecast volcanic activity^{9–13}. Our results show how the combination and correlation of the two methods can be used to interpret the movement of gases from the magma body to the surface, and to monitor the pressure buildup leading to an eruption at explosive volcanoes.

Galeras volcano (4,200 m high), a recent active cone in southern Colombia, 7 km west of Pasto (population of >300,000) (Fig. 1), was chosen as South America's 'Decade volcano', to be studied extensively during the International Decade of Natural Disaster Reduction. Galeras showed signs of unrest in early 1988, confirmed by gas sampling¹⁴, and is being regularly monitored with six seismic stations, two electronic inclinometers and gas sampling of the fumaroles. The SO₂ flux from the volcano has been monitored remotely, since 1989, at a distance of >5 km from the crater with the correlation spectrometer (COSPEC)¹⁵.

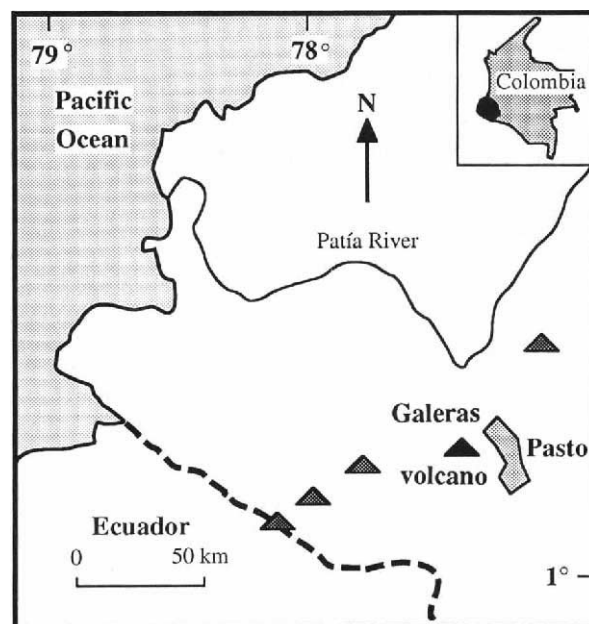


FIG. 1 Location of Galeras volcano in Colombia. Other active volcanoes are shown by triangular symbols.

A minor magmatic eruption in May 1989 caused fear of potential pyroclastic flows, recognized as typical of many prehistoric eruptions^{16,17}. SO₂ emissions reached a peak of 5,000 metric tons per day (t d⁻¹) in late 1989 (ref. 14). In October 1991, a lava dome was emplaced within the inner crater, accompanied by another peak in SO₂ flux, vulcanian explosions and a large number of seismic events per day¹⁸. Approximately 90% of this dome was explosively destroyed on 16 July 1992 (ref. 19).

Long-period seismicity occurred before the 14 January 1993 and 23 March 1993 eruptions. Shallow long-period events are commonly recognized by a high-frequency onset superimposed on a low-frequency (1–5 Hz) monotonic wave form^{20–22}. The spectra are sharply peaked and the dominant frequencies have not been observed to vary with the location of the seismic station. This suggests that the low dominant frequencies are caused by the source of the event, and not by the paths travelled by the signal from the source to the seismic stations. From these observations, the source of long-period spectrum has been modelled as the resonant modes in a fluid-filled crack excited by a localized pressure fluctuation^{5,6}. In this fluid-filled crack model, the magnitude of the pressure fluctuation is reflected in the amplitude and duration of the seismic signal. The origin of the pressure fluctuation may be bubble bursts in a vesiculating fluid²³, occurrence of a shock wave in the flow²⁴, or unsteady fluid flow⁷ (for example, release of magmatic gases).

Seventeen long-period events occurred between 1 January 1993 and the 14 January eruption. Durations and amplitudes of the long-period events continuously increased, for the first 11 days, then decreased up to eruption (Fig. 2). The number of the long-period events increased substantially for 24 h after the eruption, with individual durations ≤20 s. Seismicity was accompanied by observed gas and steam emission. After 24 h, the seismicity declined to levels lower than before the eruption²⁵.

Before the 23 March eruption, seismic activity was dominated by long-period events, with peak frequencies between 2.35 and 4.00 Hz, located west of the active crater at depths between 0.2 and 1.0 km (ref. 2). Until 13 February, only one long-period event of duration >20 s occurred. From 14 February, to until the eruption on 23 March, durations of long-period events increased to 185 s on 12 March then decreased to 45 s on 22 March (Fig. 2). Amplitude of events generally show the same

trend. For 10 h before the eruption, no long-period events of significant duration and amplitude occurred.

The SO_2 flux measurements made in January, 48 h after the fatal eruption, showed a decline from 90 t d^{-1} on 16 January to 20 t d^{-1} on 18 January. SO_2 flux increased to 560 t d^{-1} on 4 February, then declined to $<50 \text{ t d}^{-1}$ from 24 February until 12 March (Fig. 2). On 23 March at 22:39 local time, an eruption larger than 14 January occurred². Eleven hours later (24 March), SO_2 flux was 880 t d^{-1} . On 26 March, this flux declined to 580 t d^{-1} , then decreased to levels measured after the July 1992 eruption (Fig. 2).

SO_2 fluxes, which we interpret as gases originating from the shallow magma body²⁶, may be compared with the long-period signals by applying the fluid-filled crack model. At Galeras, we attribute the unsteady release of magmatic gases through cracks (between 16 January and 28 March, 1993) to the origin of pressure fluctuations causing the long-period seismic events. This allows us to make inferences about the movement of gases from the magma to the surface and on the pressurization of the volcanic edifice, as the magnitude of pressure fluctuations in the fluid is reflected in the amplitude and duration of long-period events. SO_2 flux data recorded after the 14 January 1993 eruption up until the eruption of 23 March 1993 are consistent with the model and suggest the pressurization of the volcanic edifice, eventually resulting in explosive eruption.

Three phases are recognized, each interpreted as reflecting changes in fluid dynamic conditions along pathways in which gases flow through the edifice (Fig. 3). The first phase (15 January to 13 February) is characterized by a very low number of seismic events, (including long-period events) and an increase in SO_2 suggesting that gases are moving freely through a dendritic system of fractures, resulting in relatively aseismic fluid-flow conditions from the magma to the surface, similar to periods of quiescence¹⁸.

The second phase (13 February to 12 March) is characterized by frequent long-period events with increasing duration and amplitude while the flux of SO_2 shows a gradual decrease to the lowest levels recorded. The long-period events may reflect unsteady release of gases in pathways which are segments in the dendritic system of fractures within the edifice. The unsteady release of gases results from the establishment of a pressure gradient across the long-period source region, between the magma body and the altered zone (Fig. 3). The establishment of this pressure gradient is due to a decrease in permeability

within the edifice caused by alteration of the edifice or rheological changes in the magma body. Alteration of the edifice occurs as a result of reactions with acidic gases and steam rising from the magma body, leading to the deposition of clay and hydrothermal minerals¹⁶. Pathways become clogged within the long-period source region, causing a build-up of pressure below. Seismic flow conditions arise when the gas pressure exceeds the tensile strength of the walls, opening the pathways for the gases to flow through. With time, hydrothermal mineralization continues and higher gas pressure is required to open the clogged pathways. Consequently, the magnitude of the pressure fluctuations increases, leading to an increase in the amplitude and duration of the observed long-period events. Pressure in the altered zone continues to increase, and gradually the pressure gradient across the long-period source region reduces, leading to phase three. The decrease in SO_2 flux in this phase reflects the decrease of permeability and clogging of the pathways, not changes in the release of gases by the magma body.

The third phase (13 to 23 March) is characterized by a decrease in the amplitude and duration of the long-period events, leading to a complete lack of long-period seismicity for 10 h before the eruption. As sealing of the system continues the pressure in the altered zone continues to rise, which causes further lowering of the pressure gradient across the long-period source region. This reduced pressure gradient inhibits upward flow of gases to the surface, and reduces the magnitude of the related pressure fluctuation. These flow conditions cause the observed decrease in amplitudes and durations of the events and the diminishing occurrence of long-period events. Although no measurements were taken, we assume that the SO_2 flux remained low during the phase due to continuous sealing of the system. The idea that gas was accumulating at depth is supported by SO_2 flux more than 10 times higher on 24 March, just after the eruption.

We have described how the occurrence of long-period events, coupled with SO_2 flux measurements, can be used to evaluate the pressurization of a volcano. This may be applied cautiously to the fatal eruption of 14 January. Figure 3 shows an almost continuous increase in amplitudes of long-period events from 23 December to 11 January. Amplitudes of remaining events leading up to the eruption show a continuous decrease. These trends may be explained with the same reasoning as the events before the 23 March eruption. After the eruption of 16 July 1992, the gases were escaping freely indicating that the permeability in the

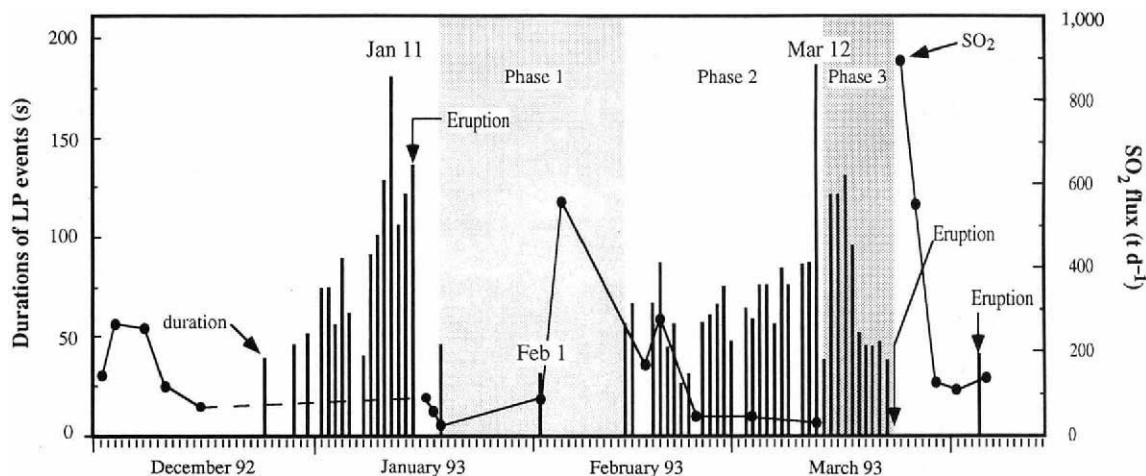


FIG. 2 SO_2 flux in metric tons per day (●) and durations of recorded long-period (≥ 22 s) events (vertical bars) plotted against time, also showing the eruptions during the same time period. The SO_2 flux is measured using correlation spectrometer (COSPEC) methodology¹⁵. Uncertainty depends mostly on recorded wind speeds. The error at

Galeras is assumed to be $\pm 20\%$, in general, and $\pm 40\%$, in the worst case. True SO_2 flux is likely to be higher than the calculated value. The three phases reflect changes in fluid dynamics along pathways through which gases flow, as interpreted from seismic and gas flux data.

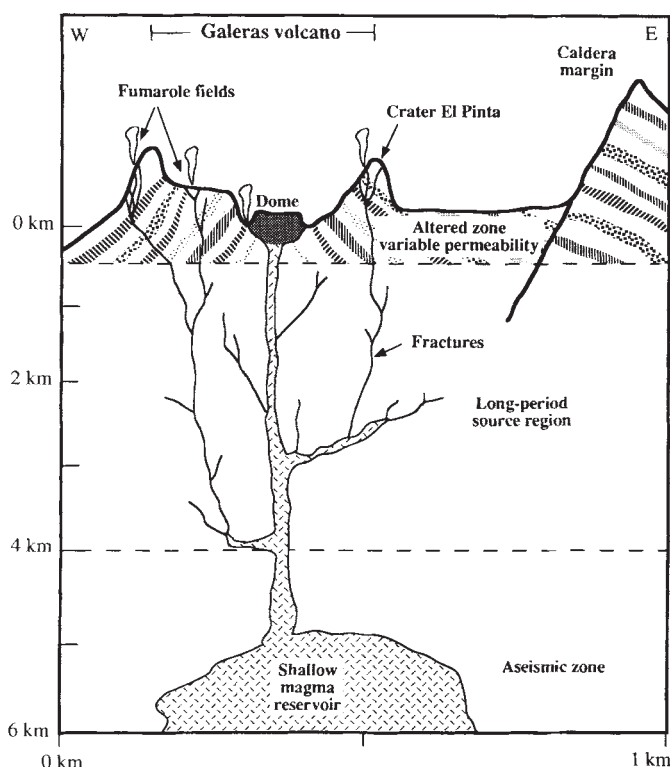


FIG. 3 Schematic model of Galeras showing zones referred to in text. The altered zone consists of highly altered material due to fumarolic activity¹⁶. The long-period source region is defined by the depths of long-period events². The aseismic zone is characterized by an overall lack of seismic activity, and is therefore assumed to be the location of the shallow magma reservoir^{16,19}.

altered zone was high. The occurrence of long-period events on 24 December 1992 may have reflected a first decrease in permeability. Observed trends in amplitudes and durations of these events show an increased pressurization, released by the fatal eruption on 14 January.

Similar trends in SO_2 and/or long-period events have been observed at various volcanos. Karkar volcano (Papua New Guinea) erupted explosively in 1979 and killed two scientists. Earlier seismic data showed an increase in volcanic tremor amplitude, followed by a clear drop for about two weeks before the most energetic explosion of the sequence⁹. Arenal volcano (Costa Rica) has had a several-week decline in tremor amplitude before eruptions¹⁰. A pattern was recognized at Mount Pinatubo volcano when SO_2 flux declined from 5,000 to 260 t d^{-1} during three weeks preceding the climactic eruption of 15 June 1991 (ref. 11). At Redoubt volcano, Alaska (1989–90 eruption) SO_2 monitoring began after the main explosions but showed a decline over three days before 23 March, 1990 eruption; the changes in degassing and long-period seismicity associated with eruptions were interpreted as the result of pressurization due to sealing of the dome^{12,13}. These examples are consistent with our model which combines gas and seismic data and suggests a sealing of the system, causing gas accumulation and increased pressurization of the volcanic edifice before explosive eruptions. □

Received 22 July 1993; accepted 14 January 1994.

1. Venke, E., Murray, T. L. & Goldberg, J. (eds) *Bull. Glob. Volcanism Network, Smithsonian Instn.* **18**(1), 31 January (1993).
2. Venke, E., Murray, T. L. & Goldberg, J. (eds) *Bull. Glob. Volcanism Network, Smithsonian Instn.* **18**(3), 13 March (1993).
3. Steinberg, G. S. & Steinberg, A. S. *J. geophys. Res.* **80**, 1600–1604 (1975).
4. Aki, K., Fehler, M. & Das, S. *J. volcan. geotherm. Res.* **2**, 259–287 (1977).
5. Chouet, B. *J. geophys. Res.* **90**, 1881–1893 (1986).
6. Chouet, B. *J. geophys. Res.* **93**, 4375–4400 (1988).
7. Chouet, B. in *Volcanic Seismology* (eds Gasparini, P., Scarpa, R. & Aki, K.) 133–156 (Springer, Berlin, 1992).
8. Seidel, D., Schick, R. & Ruscetti M. *Bull. volcan.* **44**–1, 42–56 (1981).

9. McKee, C. O. et al. in *Cooke-Ravian Volume of Volcanological Papers* (ed. Johnson, R. W.) 63–84 (Geol. Surv. of Papua New Guinea Mem. No. 10, Port Moresby, 1981).
10. Barquero, R. et al. in *Volcanic Seismology* (eds Gasparini, P., Scarpa, R. & Aki, K.) 84–95 (Springer, Berlin, 1992).
11. Daag, A. S. et al. (abstr.) *EOS* **71**, 61 (1991).
12. Casadevall, T. J. et al. *J. Volcan. geotherm. Res.* (in the press).
13. Chouet, B. et al. *J. Volcan. geotherm. Res.* (in the press).
14. Williams, S. N. et al. (abstr.) *EOS* **71**, 647 (1990).
15. Stoiber, R. E., Malinconico, L. L. & Williams, S. N. in *Forecasting Volcanic Events* (eds Tazieff, H. & Sabroux, J.-C.) 425–444 (Elsevier, Amsterdam, 1983).
16. Calvache V., M. L. thesis, Louisiana State Univ. Baton Rouge (1989).
17. Calvache V., M. L. & Williams, S. N. *Geology* **20**, 539–542 (1992).
18. Stix, J. et al. *Geology* **21**, 963–967 (1993).
19. Gomez, M. D. & Torres, C. R. *Galeras Workshop* (abstr.) Observatorio Volcanológico del Sur, 9–15 Jan, Pasto, Colombia (1993).
20. Minakami, T. *Bull. Earthq. Res. Inst.* **27**, 123–128 (1949).
21. Latter, J. H. N. *Z. Dep. Sci. Ind. Res.* **150**, 60 (1979).
22. Fehler, M. & Chouet, B. *Geophys. Res. Lett.* **9**, 1017–1020 (1982).
23. Kieffer, S. W. *J. volcan. geotherm. Res.* **22**, 59–95 (1984).
24. Martinelli, B. *Episodes* **14**, 19–25 (1991).
25. Muñoz, et al. *EOS* **74**, 281–287 (1993).
26. Archart, G. B., Sturchio, N. C., Fischer, T. C. & Williams, S. N. *Geol. Soc. Am. Abstr. Program* **25**, A-326–327 (1993).

ACKNOWLEDGEMENTS. We thank G. P. Cortéz and A. Charland for their COSPEC data, the staff of Observatorio Volcanológico del Sur for their assistance, T. Casadevall and B. Chouet for a review and the US NSF, Canadian NSERC and INGEOINAS for support. This work is dedicated to the memory of our colleagues killed by the 14 January 1993 eruption.

Biological compasses and the coordinate frame of landmark memories in honeybees

T. S. Collett & J. Baron

Sussex Centre for Neuroscience, School of Biological Sciences, University of Sussex, Brighton BN1 9QG, UK

MANY hymenopterans use visual landmarks to guide the last stages of their return to a familiar place, moving so that the pattern of landmarks imaged on their retina matches the pattern stored on previous visits to that place^{1,2}. What is the coordinate frame of these landmark memories, and how is it established? On the one hand, bees and flies learn complex visual shapes retinotopically^{3,4}, and landmark memories probably share this characteristic. On the other hand, bees record the position of landmarks in compass coordinates. Thus, Lindauer⁵ showed that bees that had been trained to feed at the southernmost corner of a square table recognized the corner by its compass bearing from the table's centre. Taken together, these results suggest that these insects place retinotopically localized memories in Earth-based coordinates. We report here that honeybees accomplish this very simply: when learning about or searching for a goal, they face consistently in one compass direction, aided by the Earth's magnetic field. We suggest that the main benefit of inspecting the world from one favoured direction is to simplify the storage and retrieval of retinotopic memories.

We found that bees have a preferred viewing direction by training them in the open air to collect sucrose from a small bottle-cap placed on a circular board, 170 cm in diameter, lying on the ground. The position of the bottle-cap on the board was marked by a black cylinder (3 cm in diameter and 16.5 cm high) placed in a constant compass direction 15 cm away. Feeder and cylinder were moved to a different location on the board after every visit. After a day's training, we periodically removed the bottle-cap and recorded on videotape an individually marked bee's course as it flew a few centimetres above the board searching for sucrose. Videotapes were later digitized to extract the bee's position and orientation in the horizontal plane.

The orientations and positions adopted during search flights were accumulated and converted into circular histograms. Orientation is specified in compass bearings and position is given radially, as the bee's compass bearing relative to the landmark. Figure 1 shows that the bee spent most time facing south and searching west of the cylinder—the direction of the sucrose.

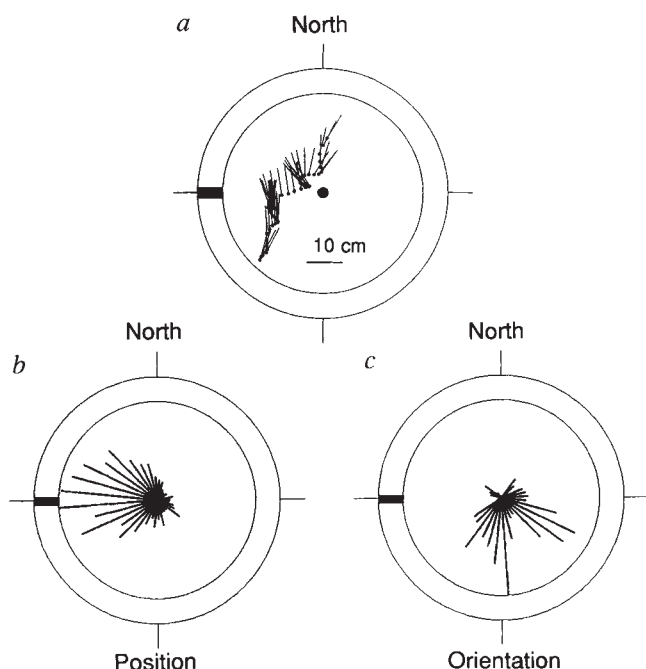


FIG. 1 Bee's position and orientation during tests in which the bottle-cap was removed. *a*, Two-second excerpt from search flight by bee trained with food placed to west of cylinder. Bee flies from top to bottom of picture. 'Tadpoles' indicate bee's position and orientation every 40 ms. Here, and in Fig. 2, filled sector in the annulus gives bearing of food with respect to cylinder. Recordings were made using a camera suspended 2.2 m above the board. *b*, Bee's radial position with respect to the landmark in 10° bins accumulated over two tests with those portions of the flight excluded in which the bee was >16 cm or <3.5 cm from the cylinder ($n=2,573$). *c*, Same bee's orientation during one test ($n=891$ frames). Orientation was only digitized when the bee was on the same side of the cylinder as the sucrose during training. Digitized frames in which the bee fixated the cylinder (that is, when its long axis pointed within $\pm 10^\circ$ of the centre of the cylinder) are excluded from the distribution. The directions and magnitudes of the mean vectors of the distributions are 275° and 0.467 (position) and 156° and 0.533 (orientation).

Twenty-eight bees were trained with the sucrose placed north, south, east or west of the cylinder. Figure 2 gives the mean vectors⁶ of each bee's position and orientation histogram and shows that, after training with food to the north, west or east, bees tend to face south.

Bees accustomed to finding sucrose to the south of the cylinder provide a revealing exception. If they adjust their position relative to a landmark while facing south, then a place that is defined by a narrow cylinder to its north will be hard to locate: when a bee looks south, such a landmark will be behind it and out of its field of view. Bees given this problem started by facing south, but after a day's training their preferred orientation shifted to the north-west, so rendering the cylinder visible (Fig. 2). This shift suggests that bees store memories of landmarks while in their preferred orientation and adopt the same viewing direction when using the memory in searching for a goal.

Pointing south is not peculiar to extended search flights. A normal approach to the feeder consists of two navigational phases. Bees arrive in the area surveyed by the camera aiming directly at the cylinder; this suggests that the cylinder, seen at a distance, is treated as an attractive beacon. Only when bees are close to the cylinder do they face southwards (Fig. 3*a*) so that the cylinder can take up its stored retinal position. To exclude the first phase of the approach from analysis, we omitted portions in which the bee faced within $\pm 10^\circ$ of the cylinder. The average

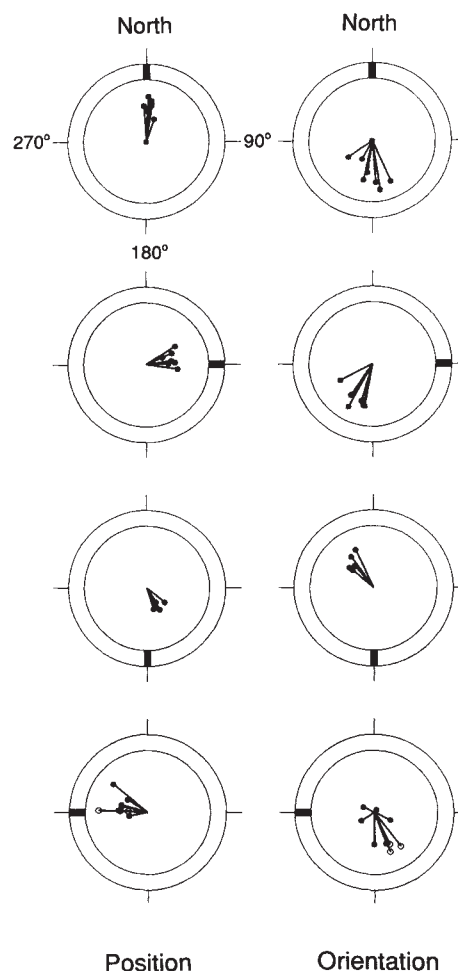


FIG. 2 Mean position and orientation vectors of 28 bees computed from circular histograms like those in Fig. 1. The radius of the inner circle represents a vector amplitude of one. Mean amplitude of position and or orientation vectors are, respectively, 0.40 and 0.44. Solid circles indicate bees trained and tested in the open; open circles denote bees trained in the tent. The second-order orientation vector of bees trained north, east or west of the cylinder has a direction of 180° and an amplitude of 0.728. Application of the Rayleigh test with a sample size of 23 gives a P value of <0.001 , indicating that the bees' southward orientation cannot be ascribed to chance. According to the Mardia-Watson-Wheeler test⁶, the distribution of orientation vectors from bees trained with the food to the south of the landmark differs significantly from that of bees trained with the food north, east or west of it ($P < 0.01$).

(second-order⁶) vector was then calculated from the mean orientation vectors of 51 digitized approaches of bees newly recruited to a feeder placed east or west of the cylinder. Vector direction was 170° and its amplitude 0.71 ($P < 0.001$ that the direction is random). The orientation of experienced bees before landing is more erratic. They may have come to recognize the bottle-cap as a secondary beacon and so could choose to fly directly towards it without facing south.

Bees and wasps memorize the landmarks near a significant place both on their approach⁷⁻⁹ and during specialized 'orientation' flights performed on their first few departures from it⁷⁻¹¹. To make these memories congruent, bees should adopt the same viewing direction on their initial departures as on arrival, as do solitary wasps when learning about and searching for their nest^{11,12}. Orientation vectors obtained from 22 orientation flights with the feeder located mostly east of the cylinder (for example, Fig. 3*b*) support this contention. Because there may be different

FIG. 3 Bee's position and orientation every 20 ms. *a*, Landing: the trained bee arrived at the bottom right of the frame, heading for the cylinder (solid circle); it faced south just before landing on the feeder (cross). *b*, Orientation flight performed on departure; the bee left the feeder and flew north in arcs of increasing radius during which time it faced south. The hive was east-south-east of the board.

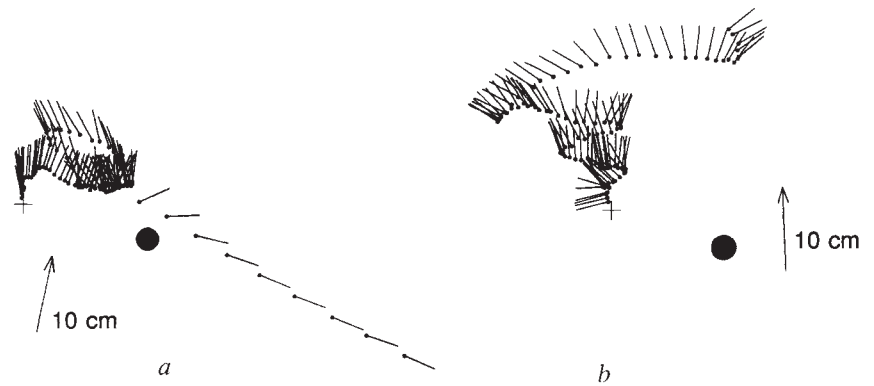
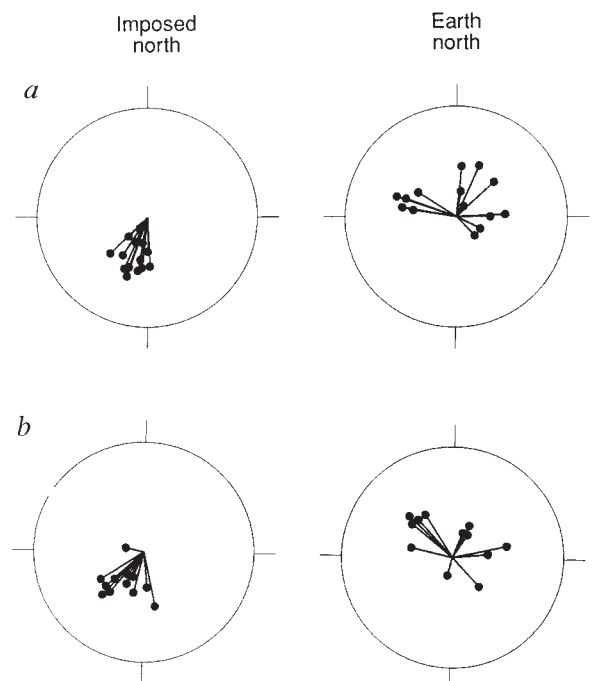


FIG. 4 Mean orientation vectors of 14 bees trained with an imposed magnetic field. *a*, Tests with magnet in place. *b*, Tests from 12 bees with the magnet removed. Two rows of permanent ferrite magnets, 67 cm apart on a steel plate, were placed 30 cm beneath the circular board. The resultant magnetic field on the board was roughly linear over a radius of about 30 cm from the cylinder with a field strength of 0.55 mT. Training was in one of three arrangements. (1) Imposed north was aligned with Earth-south and the feeder was Earth-west of the cylinder. (2) Imposed north was aligned with Earth-west and the feeder was Earth-north of the cylinder. (3) Imposed north was aligned with Earth-east and the feeder was Earth-east of the cylinder. Because the bee tends not to face away from the cylinder (Fig. 2), this last condition can only show that the imposed field controls the bee's axis of orientation. Orientation is shown with respect to the imposed magnetic field (left) and the Earth's field (right). The second-order vector for tests with magnets is 200° , with an amplitude of 0.965. In the absence of magnets, this vector is 214° , with an amplitude of 0.88. Some of the small-amplitude vectors are from bees which alternated between heading imposed south and north. In six cases the amplitude of the bimodal vector, obtained by doubling angles⁶, was the same amplitude or larger than the unimodal vector.



viewing strategies for learning the appearance of the cylinder and the location of the feeder, separate vectors were computed for episodes in which bees faced within $\pm 10^\circ$ of the feeder or the cylinder. The average of the 22 mean vectors was significantly oriented in both cases ($P < 0.01$), and it was significantly ($P = 0.026$) closer to south (173° , 0.53 amplitude) during fixation of the feeder, an appropriate time to store the bearing of the feeder, than it was during fixation of the cylinder (146° , 0.62 amplitude). It is unclear why our findings differ from those of Lehrer⁹, who saw no directional pattern in the orientation flights of bees leaving a feeding dish.

Several directional cues¹³ probably contribute to the bees' southward orientation. The 'sky compass'¹⁴ alone does not suffice, for bees faced south in the rain under a completely overcast sky. They did the same when we eliminated the sky compass with a tent constructed out of a double layer of white tarpaulin, which depolarized skylight and filtered out ultraviolet light. Figure 2 (open circles) shows the orientation vectors of three bees trained in the tent with the food placed to the west.

Direct evidence for the involvement of a magnetic sense¹⁵⁻²⁰ was obtained from training bees under the tarpaulin in an imposed magnetic field directed to shift magnetic north 90° or 180° from Earth-north. The bees' preferred orientation during

search flights was south with respect to the imposed magnetic field (Fig. 4*a*, left). When the same data were replotted in terrestrial coordinates, the directions of the mean vectors were scattered and, in contrast to those of Fig. 2, did not point south (Fig. 4*a*, right).

When the same bees were tested with the magnets removed, they did not revert to facing Earth-south, but persisted in the direction that had been dictated by the bar magnets (Fig. 4*b*). The insects must have memorized views of both the cylinder and the surrounding panorama in the tent while facing south in the imposed field. Without the magnets, they turned until the panorama took up its accustomed retinal position. The bees' viewing direction was thus controlled by the panorama rather than by conflicting signals from the Earth's magnetic field. It is not known whether magnetic compass cues influence orientation in a homogeneous landscape⁵.

Bees adopt different viewing directions under different circumstances. When newly departing bees learn the location of the hive entrance, they back away, looking at the entrance and pointing so that the Sun lies behind them²¹. On the other hand, experienced bees, caught while flying out of a hive with its entrance facing east, continued in an easterly direction when released in an arena in the dark within an imposed magnetic field²⁰. Perhaps bees on their outward journeys learn about and

search for foraging sites while facing in the direction in which they leave the hive, whereas bees learning how to return home face towards the entrance.

In conclusion, the bees we studied tended to face south. They did this both while learning about the immediate vicinity of a feeder and when returning to it. In this way, retinotopic and compass-based coordinates are automatically kept in register. Two findings suggest that a standard viewing direction is impor-

tant in handling retinotopic memories. First, the preferred orientation shifts away from south when the defining landmark lies to the north of the feeder and so cannot be seen by a bee facing south; and second, bees hold to the shifted orientation induced by an imposed magnetic field after the magnets have been removed. Bees can thus orientate by means of their visual surroundings using memories stored while facing magnetic south. □

Received 12 October 1993; accepted 18 January 1994.

1. Wehner, R. in *Handbook of Sensory Physiology* vol. VII/6C (ed. Autrum, H. J.) 287–616 (Springer, Berlin, 1981).
2. Collett, T. S. *Phil. Trans. R. Soc. Lond. B* **337**, 295–303 (1992).
3. Wehner, R. *J. comp. Physiol.* **77**, 256–277 (1972).
4. Dill, M., Wolf, R. & Heisenberg, M. *Nature* **365**, 751–753 (1993).
5. Lindauer, M. *Cold Spring Harb. Symp. quant. Biol.* **25**, 371–377 (1960).
6. Batschelet, E. *Circular Statistics in Biology* (Academic, London, 1981).
7. Opfinger, E. Z. *vergl. Physiol.* **15**, 431–487 (1931).
8. Couvillon, P. A. & Bitterman, M. E. *J. Insect Behav.* **5**, 123–129 (1992).
9. Lehrer, M. *J. comp. Physiol.* **A172**, 549–563 (1993).
10. Tinbergen, N. Z. *vergl. Physiol.* **16**, 305–335 (1992).
11. Zeil, J. *J. comp. Physiol.* **A172**, 189–205 (1993).
12. Zeil, J. *J. comp. Physiol.* **A172**, 207–222 (1993).

13. von Frisch, K. *The Dance Language and Orientation of Bees* (Belknap, Cambridge, 1967).
14. Wehner, R. & Rosell, S. in *Experimental Behavioral Ecology and Sociobiology* (eds Hölldobler, B. & Lindauer, M.) 11–53 (Sinauer, Sunderland, 1985).
15. Lindauer, M. & Martin, H. Z. *vergl. Physiol.* **60**, 219–243 (1968).
16. Martin, H. & Lindauer, M. *J. comp. Physiol.* **122**, 145–187 (1977).
17. Gould, J. L., Kirschvink, J. L. & Deffeyes, K. S. *Science* **201**, 1026–1028 (1978).
18. Walker, M. M. & Bitterman, M. E. *J. exp. Biol.* **145**, 489–494 (1989).
19. Kirschvink, J. L. & Kobayashi-Kirschvink, A. *Am. Zool.* **31**, suppl. (1), 169–185 (1991).
20. Schmitt, D. F. & Esch, H. E. *Naturwissenschaften* **80**, 41–43 (1993).
21. Vollbrecht, J. *Zool. Jb. Physiol.* **79**, 33–69 (1975).

ACKNOWLEDGMENTS. We thank E. Godwin for the loan of bees, and M. Land and D. Osorio for scrutinizing the typescript. Financial support came from SERC.

Turning of nerve growth cones induced by neurotransmitters

James Q. Zheng, Mark Felder, John A. Connor* & Mu-ming Poo†

Department of Biological Sciences, Columbia University, New York, New York 10027, USA

* Roche Institute of Molecular Biology, Roche Research Center, Nutley, New Jersey 07110, USA

PATHFINDING by growing nerve processes in the developing nervous system depends on the turning response of the growing tip, the growth cone, to extracellular guidance cues^{1–4}. There is evidence *in vivo* and in cell culture that some growth cones exhibit chemotactic behaviour^{5–12}, but the identity of endogenous chemoattractants remains elusive. Neurotransmitters appear early in the developing embryo and may have morphogenic roles in development^{13,14}. In cell culture a number of neurotransmitters were found to induce growth inhibition or retraction of neurites^{15–19}. Here we report positive turning responses of the nerve growth cone in a defined extracellular gradient of the neurotransmitter acetylcholine (ACh). The growth cone response depends on the activation of neuronal nicotinic ACh receptors, requires the presence of extracellular Ca^{2+} , and appears to be mediated by Ca^{2+} -calmodulin-dependent protein kinase II. Fluorescence imaging of cytosolic Ca^{2+} concentration ($[Ca^{2+}]_i$) at the growth cone showed a small but significant elevation of $[Ca^{2+}]_i$ within minutes of the onset of ACh application and before the turning of the growth cone. These findings suggest that neurotransmitters may serve as specific chemoattractants for growth cone guidance and that cytosolic Ca^{2+} may act as a second messenger in the cytoplasm of the growth cone to initiate the turning response.

Isolated embryonic spinal neurons in *Xenopus* cell cultures were used. Microscopic gradients of ACh molecules near the growth cone were established by repetitive pulse application of picolitres of concentrated solution from a micropipette²⁰. The profile of chemical gradients produced by this method was monitored quantitatively using a fluorescent dye, 5-carboxyfluorescein. Under repetitive pulse application, a relatively stable gradient of chemicals can be quickly established by diffusion over a distance of a few hundred micrometres (Fig. 1). Quantitative measurement indicates that at a distance of 100 μ m from the source of the chemical, a gradient of up to ~12% can be

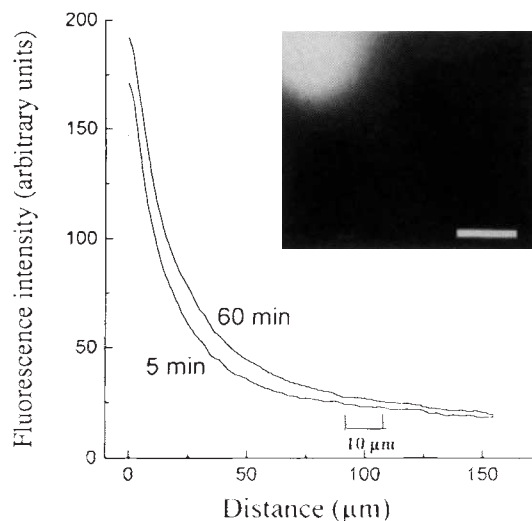


FIG. 1 Gradient of chemicals created by repetitive pulsatile ejection from a micropipette. 5-carboxyfluorescein was used to demonstrate the gradient. The fluorescence intensity surrounding the source pipette was measured 5 and 60 min after the onset of repetitive pulses. The steepness of the gradient at 100 μ m (across 10 μ m distance, as marked) was estimated to be 4 and 12%, respectively, at 5 min and 60 min. A fluorescent image of the gradient (at 60 min) is shown in the inset (scale bar, 50 μ m).

METHODS. Micropipette (tip opening, 1 μ m) was filled with 1 mM 5-carboxyfluorescein dissolved in culture medium, and was connected to an electrically gated pressure application system (Picospritzer, General Valve Co.). Positive pressure of 3 psi was applied to the pipette with a frequency of 2 Hz and a duration of 10 ms using a pulse generator (SD9, Grass Co.). The average volume of the ejected solution per pulse was estimated to be in the range 0.3–0.7 pl (ref. 20). Pulsatile ejection of 5-carboxyfluorescein and fluorescence microscopy were performed on an inverted microscope (Axiovert 35, Zeiss) equipped with fluorescence attachments and a $\times 20$ objective. After the onset of pulsatile ejection, fluorescent images were collected by a cooled CCD (charge coupled device) camera (Star I, Photometrics, Tucson, AZ), which has a dynamic range of up to 4096 gray levels in intensity. The intensity profile of the fluorescent gradient was then measured pixel by pixel over a distance of a few hundred microns. The background reading before application of fluorescence dye was subtracted from all intensity measurements.

† To whom the correspondence should be addressed.

TABLE 1 Response of *Xenopus* nerve growth cones in ACh gradients

[ACh] in pipette	Chemicals in bath	Turning angle (deg)*	Net neurite extension (μm)*	Turning responses (%)†			Number of cells examined
				+	0	-	
0 mM	None	-5.1 ± 4.6	24.9 ± 2.1	33	33	33	21
10 mM	None	-1.2 ± 5.5	30.8 ± 3.6	35	35	30	23
30 mM	None	$9.4 \pm 5.4^{**}$	23.0 ± 2.7	62	19	19	21
100 mM	None	$21.0 \pm 3.8^{**}$	25.2 ± 2.9	71	17	12	41
100 mM	D-Tubocurarine (100 μM)	-2.9 ± 5.6	25.3 ± 2.6	41	16	43	37
100 mM	Atropine (1 μM)	$10.1 \pm 5.2^{**}$	$36.9 \pm 4.0^{**}$	50	15	35	20
100 mM	Calcium-free‡	-7.7 ± 4.6	$84.2 \pm 3.8^{**}$	17	22	61	18
100 mM	KN62 (2.5 μM)§	-8.0 ± 6.1	24.3 ± 3.5	37	21	42	24

* Pulses of ACh were pressure-ejected into the bath through a micropipette at a position 100 μm away from the growth cone and at an angle of 45° with respect to the direction of neurite extension. The turning angles with respect to the original direction of growth were determined by a straight line connecting the position of the growth cone at 2 h after the onset of ACh application to that at the onset. The length of neurite extension was measured for the entire trajectory of the path of extension during the 2-h period (Fig. 2e, f). Values represent mean $\pm$ s.e., Values marked with double asterisks were different from the control (first row, $P < 0.05$, t -test). The value of the turning angle for the atropine experiment was not significantly different from that of ACh (100 mM) or ACh (30 mM) with $P > 0.05$.

† Growth cone turning responses were scored as follows: plus sign refers to the percentage of cells showing positive turning towards the ACh source (turning angle $> 5^\circ$); 'zero' refers to cells showing no turning ($|\text{turning angle}| \leq 5^\circ$); and minus sign refers to cells turning away from the ACh source (turning angle $< -5^\circ$).

‡ Calcium-free solution: 115 mM NaCl, 2.6 mM KCl, 5 mM MgCl_2 , 2 mM EGTA, 10 mM HEPES, pH 7.4.

§ KN62, which is 1-[N, O-bis(5-isoquinolinesulphonyl)-N-methyl-L-tyrosyl]-4-phenylpiperazine (Calbiochem), was added to the bath solution 20 min before ACh application and was present through the experiment.

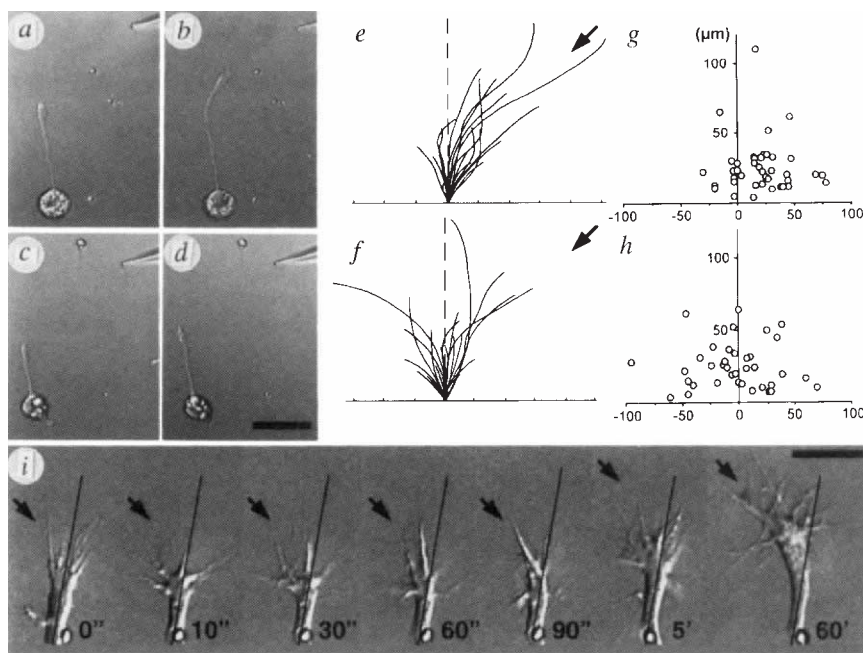
established across a distance of 10 μm , or approximately the width of a growth cone, a result consistent with a previous theoretical estimate²⁰. In this study, the source of ACh was placed 100 μm from the centre of the growth cone and at an angle of 45° from the direction of neurite extension. With an ACh concentration of 100 mM in the source pipette (tip opening, 1 μm) and standard pulse parameters (frequency, 2 Hz; dura-

tion, 10 ms; pulse pressure, 3 psi), consistent turning of the growth cone towards the pipette tip was observed within 2 hours of the onset of ACh application (Fig. 2a, b). No turning was observed when D-tubocurarine, a nicotinic antagonist, was present in the bath medium (Fig. 2c, d). To illustrate better the ACh-induced turning response, the path of neurite extension of a number of cells during the 2-h period of ACh application in

FIG. 2 Turning response of *Xenopus* spinal neurons in the presence of ACh gradient. The gradient was created by pulsatile ejection of picolitres of 100 mM ACh solution through a micropipette. a–d, Representative DIC images at the onset and 2-h after the application of ACh gradient in the absence (a, b) and presence (c, d) of 0.1 mM D-tubocurarine. Scale in d is 50 μm . e, f, Composite drawings of the path of neurite extension during a 2-h period for a population of neurons in the absence (e) and presence (f) of D-tubocurarine (0.1 mM). For clarity, only part of the data (all cases examined during the most recent 8-month period) were illustrated for each drawing. The origin represents the position of the centre of the growth cone palm at the onset of the ACh application. The original direction of neurite extension, as defined by the last 20- μm segment of the neurite at the onset of the ACh application, was aligned with the vertical dashed line. The line depicts the trajectory of the neurite after a 2-h exposure to the ACh gradient. Arrows indicate the direction of the ACh gradient. Scale, 10 μm . g, h, Scatter plots of all data corresponding to experiments illustrated in a–f. Each point represents the result of the turning angle and the net neurite extension, of one 2-h ACh gradient experiment in the absence (g) and presence (h) of D-tubocurarine (abscissa in degrees, ordinate in μm). i, DIC images of a growth cone subjected to the ACh gradient before and at various times after the onset of the ACh gradient. Arrows indicate the direction of the ACh pipette. Lines indicate the original direction of neurite extension. Note the asymmetric filopodial distribution within the first minute after the onset of ACh application, and the eventual turning of the growth cone at 60 min. Scale, 20 μm .

METHODS. Cultures were prepared according to procedures reported previously^{38–40}. Briefly, the neural tube tissue from 1-day-old embryo was dissociated in Ca^{2+} - and Mg^{2+} -free Ringer's solution supplemented with EDTA, plated on clean glass coverslips, and incubated at room temperature (20–22 $^\circ\text{C}$) for 6 to 10 h before the experiments. The cul-

ture medium consists of 50% (v/v) of Leibovitz medium (GIBCO), 2% (v/v) of fetal calf serum (GIBCO), and 48% (v/v) of Ringer's solution (115 mM NaCl, 2 mM CaCl_2 , 2.5 mM KCl and 10 mM HEPES, pH 7.3). The ACh gradient was produced with a protocol similar to that described in Fig. 1. The micropipette containing 100 mM ACh chloride was positioned at a distance of 100 μm from the centre of the growth cone and at an angle of 45° from the direction of neurite extension. ACh was ejected at a frequency of 2 Hz and a pulse duration of 10 ms. The microscopic images were obtained with a 20 $\times$ objective, collected by a CCD camera (Fig. 1), processed with an image processor (Image Technology Model 151), and stored on an optical memory disk recorder (Panasonic TQ 2026F).

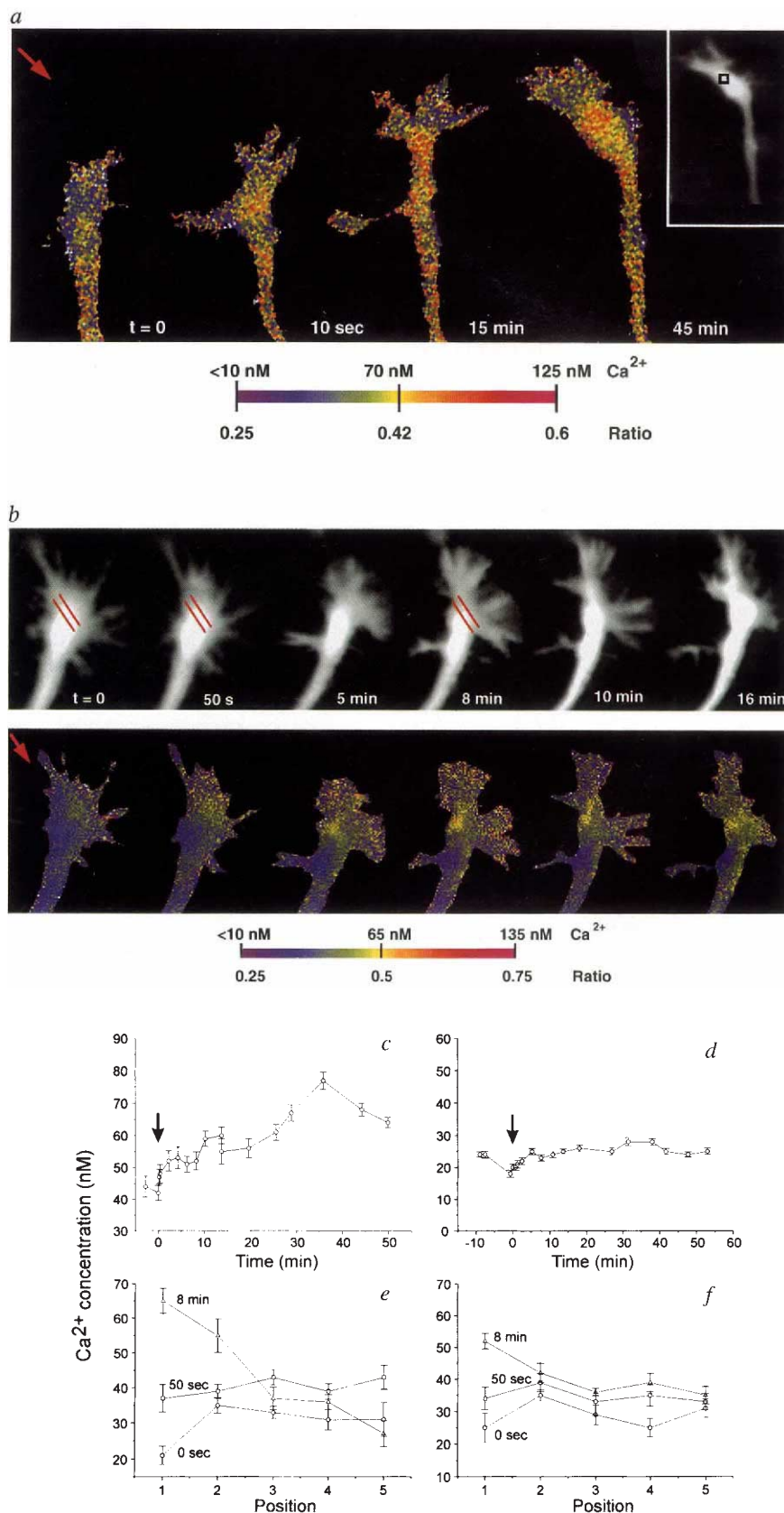


the absence and presence of D-tubocurarine in the bath is presented as composite drawings (Fig. 2e, f, respectively). The angle of growth cone turning was defined by the angle between the original direction of extension and the line connecting the position of the growth cone at the onset and at 2 h after the onset of the ACh application. Figure 2g, h depicts the data on

the turning angle and the length of extension for all growth cones examined in the experiments shown in Fig. 2a-f. Table 1 summarizes quantitative measurements of the angle of turning, the length of neurite extension, and the percentage of cells showing turning responses under various experimental conditions. Significant turning was also observed when the ACh concentra-

FIG. 3 Fura-2 imaging of $[Ca^{2+}]_i$ in the growth cone during ACh-induced turning. **a**, A growth cone showing a Ca^{2+} increase after the onset of ACh application. The level of $[Ca^{2+}]_i$ is shown in pseudocolours with the corresponding Ca^{2+} concentration and fluorescence ratio at 340 and 380 nm depicted by the colour bar. The ACh gradient was applied in the same way as for Fig. 2. The source of ACh was located 45° from the direction of neurite extension in the upper-left corner (red arrow). Fluorescence image at 380 nm excitation is shown in the inset on the right; black box shows the point of $[Ca^{2+}]_i$ measurement. **b**, Another growth cone exhibiting a gradient of Ca^{2+} during the first few minutes after application of the ACh gradient. Top panel shows a time-lapse sequence of fluorescence images of the growth cone at 380 nm excitation, and the bottom panel shows the corresponding pseudocolour Ca^{2+} images. Note the increase in the Ca^{2+} level on the side of the growth cone facing the source of ACh. Red lines depict the positions of quantitative Ca^{2+} measurements illustrated in **e** and **f**. **c**, $[Ca^{2+}]_i$ levels at different times after the onset of the ACh application at the centre of the growth cone shown in **a**. The size of the measurement box was 5×5 pixels and the data shown are mean $\pm$ s.e. Arrow indicates the onset of ACh application. **d**, Similar measurements as **c** for another growth cone in the presence of D-tubocurarine (0.1 mM) in the bath. **e** and **f**, $[Ca^{2+}]_i$ was measured along the direction of the gradient at three different times ($t=0$, 50 s and 8 min) after the onset of ACh application. Five measurements were made along the red line shown in **b** from the upper left to the lower right (corresponding to values at 1 to 5). Data in **e** and **f** are for measurements along the left and right red lines, respectively. The size of the measurement box was 3×3 pixels and the data shown are mean $\pm$ s.e.

METHODS. For fluorescence imaging of $[Ca^{2+}]_i$, *Xenopus* cultures plated on glass coverslips were incubated in culture medium containing 3–6 μ M Fura-2-AM for 30 min at room temperature. They were then carefully rinsed and mounted on the stage of an inverted microscope, equipped with a cooled CCD-based imaging system³⁴. To limit excitation beam intensity, light from a 100W Hg bulb was collected using a standard glass condenser (instead of quartz) and then passed through a 1.0 neutral density filter. Excitation wavelengths were 340/380 nm, with an exposure time of 100 ms at each wavelength. To check for de-esterification and activation of the Fura-2, ionomycin (2 μ M) was added to the culture medium at the end of experiments. Ratio values within 15% of the *in vitro* R_{max} (340/380 excited fluorescence) were routinely observed, which was indicative of nearly complete de-esterification.



tion in the ejection pipette was 30 mM. Based on a previous analysis²⁰, we estimated the ACh concentration at the growth cone to be about 1,000-fold lower, or 30 μ M for the latter condition. The presence of D-tubocurarine (100 μ M) abolished the turning response without significantly affecting the rate of extension. In contrast, significant turning was still observed when a muscarinic antagonist, atropine (1 μ M), was present in the culture medium. These findings suggest that the turning response of the growth cone in the presence of an ACh gradient was a result of the activation of nicotinic ACh channels in the neuronal membrane. Finally, neurite retraction or growth inhibition was consistently observed when glutamate (100 mM) was used instead of ACh in the ejection pipette, a result consistent with a previous report¹⁸.

There was substantial variability in the ACh-induced turning response of growth cones (Fig. 2e). This may be due in part to the heterogeneity of neuronal types in this culture, although previous studies have shown that the majority of these neurons are cholinergic in nature and are capable of forming functional synapses with co-cultured muscle cells^{21,22}. It is also possible that all growth cones are affected by the ACh gradient, but random directional biases of the growth cones before the experiment may have affected the turning response to varying degrees in different cells. Such individual variability and the overall response were also observed previously for the turning response of these neurons induced by focal electric fields²³ and by gradients of dibutyryl-cAMP²⁰.

The turning of the growth cone in the presence of the ACh gradient provides an opportunity to examine the cellular responses underlying the turning behaviour, particularly events that precede turning. High-resolution differential interference contrast (DIC) optics were used to monitor the filopodial activity of the growth cone after the onset of ACh application. As shown in Fig. 2i, more filopodia appeared on the side of the growth cone facing the gradient within minutes of the onset of ACh application but before the bending of the main trunk of the growth cone. This suggests that the initial response of the growth cone in an ACh gradient was a preferential filopodial extension towards the source of ACh.

Cytosolic Ca^{2+} levels modulate the extension and motility of growth cones in a variety of neurons²⁴⁻²⁷, and ACh triggers Ca^{2+} influx through nicotinic ACh receptors^{28,29}. Extracellular focal application of electric fields at the growth cone resulted in a local elevation of $[\text{Ca}^{2+}]_i$ and increased the number of filopodial protrusions^{26,27}. The possibility that Ca^{2+} influx is responsible for initiating the ACh-induced turning response of the growth cone was tested by experiments using Ca^{2+} -free medium. Consistent with a previous report³⁰, the rate of neurite extension was markedly increased when the culture medium was substituted with Ca^{2+} -free Ringer's saline supplemented with 2 mM EGTA and 5 mM Mg^{2+} . However, this Ca^{2+} -free medium did not deplete cytoplasmic Ca^{2+} levels (J.Z., M.-m.P. and J.A.C., manuscript in preparation). When an ACh gradient similar to those used for Fig. 2 was established near the growth cone in the Ca^{2+} -free medium, the turning response of the growth cone was abolished (Table 1), suggesting that Ca^{2+} influx is required. Furthermore, when KN62, a specific inhibitor^{31,32} of calmodulin-dependent protein kinase II (CaM kinase II), was added to the culture medium before and during application of the ACh gradient, the turning response was completely inhibited whereas the neurite extension was not affected (Table 1). Taken together with the results of experiments using channel antagonists, these findings suggest that extracellular ACh gradients act by activating nicotinic ACh channels, and that the resultant Ca^{2+} influx and subsequent activation of CaM kinase II are responsible for initiating the turning response.

Growth cone Ca^{2+} levels were monitored by ratio imaging of Fura-2 (refs 33-35). We observed a small but detectable elevation of $[\text{Ca}^{2+}]_i$ after the onset of ACh application (Fig. 3). The elevation was mainly restricted to the growth cone and occurred

before detectable turning of the growth cone. A rapid increase in $[\text{Ca}^{2+}]_i$ at the growth cone within seconds following the onset of the ACh gradient was observed in many cases (Fig. 3a, c). In the absence of ACh in the ejection pipette or in the presence of D-tubocurarine in the bath, Ca^{2+} level was stable within ~ 5 nM (Fig. 3d). We therefore consider a $[\text{Ca}^{2+}]_i$ rise of greater than 10 nM as significant. Within 30 min after the onset of the ACh gradient (100 mM in the pipette), a significant $[\text{Ca}^{2+}]_i$ increase at the centre of the growth cone was observed in 17 out of 53 growth cones examined. This may be an underestimate of the number of the neurons that actually exhibited $[\text{Ca}^{2+}]_i$ increases because of the small magnitude or the possible transient nature of $[\text{Ca}^{2+}]_i$ increase in these growth cones. The mean $[\text{Ca}^{2+}]_i$ increase at the centre of the growth cone for all cases showing significant $[\text{Ca}^{2+}]_i$ elevation (measured with box size of 5×5 pixels) was 18 ± 3 nM (s.e., $n = 17$). When there was no ACh in the ejection medium ($n = 3$) or when D-tubocurarine was added in the bath ($n = 7$), there was no increase of $[\text{Ca}^{2+}]_i$ (Fig. 3d). Furthermore, a gradient of $[\text{Ca}^{2+}]_i$ at various regions of the growth cone along the direction of the ACh gradient was found in a few relatively large growth cones (Fig. 3b, e, f). Such a cytoplasmic $[\text{Ca}^{2+}]_i$ gradient may be responsible for initiating the turning response. We noted, however, that in these imaging experiments the frequency of ACh-induced turning responses (defined as turning angle $> 5^\circ$) at the end of a 2-h period (7 out of 20 cases) was much lower than that observed without fluorescence imaging (29 out of 41 cases). Fura-2 loading and fluorescence imaging may have produced cytotoxic effects and $[\text{Ca}^{2+}]_i$ gradients may have been smeared by an indicator 'shuttling' effect. Our findings on the $[\text{Ca}^{2+}]_i$ elevation before the growth cone turning are consistent with the role of $[\text{Ca}^{2+}]_i$ in regulating growth cone motility²⁵⁻²⁷. A small increase in $[\text{Ca}^{2+}]_i$ is appropriate and required for the extension and turning response of the growth cone. Higher elevation of $[\text{Ca}^{2+}]_i$ is likely to induce growth inhibition or neurite retraction. Indeed, increases in $[\text{Ca}^{2+}]_i$ greater than 300 nM were observed following application of a glutamate gradient (data not shown), which induced retraction of these *Xenopus* neurites. Previous studies on snail and hippocampal neurons have also demonstrated a correlation between growth inhibition and a marked elevation of $[\text{Ca}^{2+}]_i$ after application of neurotransmitters or electrical stimulation^{25,26,36,37}.

In conclusion, our results indicate that a gradient of ACh may produce a positive chemotropic effect on nerve growth. It is possible that localized secretion of neurotransmitter by the target tissue or other neurons in the embryo may influence the direction of neurite growth. Whether appropriate temporal and spatial patterns of extracellular neurotransmitter exist within the developing embryo and play a role in the pathfinding of nerve processes remains to be investigated. As a gradient of $[\text{Ca}^{2+}]_i$ elevation may initiate the growth cone turning response, other diffusible or substrate-bound factors that result in an asymmetric Ca^{2+} influx at the growth cone are all potential chemoattractants. □

Received 26 July; accepted 24 December 1993.

- Purves, D. & Lichtman, J. W. *Principles of Neural Development* (Sinauer, Sunderland, MA, 1985).
- Letourneau, P. in *Molecular Bases of Neural Development* (eds Edelman, G. M., Gall, W. E. & Cowan, W. M.) 269-293 (Wiley, New York, 1985).
- Landmesser, L. *Trends Neurosci.* **9**, 489-491 (1986).
- Bray, D. & Hollenbeck, P. J. *A. Rev. Cell Biol.* **4**, 43-61 (1988).
- Letourneau, P. C. *Devel Biol.* **66**, 183-196 (1978).
- Gundersen, R. & Barret, J. N. *Science* **206**, 1079-1080 (1979).
- Lumsden, A. G. S. & Davis, A. M. *Nature* **306**, 786-788 (1983).
- Tessier-Lavigne, M., Placzek, M., Lumsden, A. G. S., Dodd, J. & Jessel, T. M. *Nature* **336**, 775-778 (1988).
- Menesini-Chen, M. G., Chen, J. S. & Levi-Montalcini, R. *Arch. Ital. Biol.* **116**, 53-84 (1978).
- Jeffner, C. D., Lumsden, A. G. S. & O'Leary, D. D. *M. Science* **247**, 217-220 (1990).
- McCraig, C. D. *J. Physiol.* **375**, 39-54 (1986).
- Yaginuma, H. & Oppenheim, R. W. *J. Neurosci.* **11**, 2598-2613 (1991).
- Buznikov, G. A., Chudakova, I. V., Berysheva, L. V. & Vyazmina, N. M. *J. Embryol. exp. Morph.* **20**, 119-128 (1968).
- Lauder, J. M. *Prog. Brain Res.* **73**, 365-387 (1988).
- Haydon, P. G., McCobb, D. P. & Kater, S. B. *Science* **226**, 561-564 (1984).

16. McCobb, D. P., Haydon, P. G. & Kater, S. B. *J. Neurosci. Res.* **19**, 19–26 (1988).
17. Lankford, K. L., DeMello, F. G. & Klein, W. L. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2839–2843 (1988).
18. Mattson, M. P., Dou, P. & Kater, S. B. *J. Neurosci.* **8**, 2087–2100 (1988).
19. Lipton, S. A., Forsch, M. P., Phillips, M. D., Tauck, D. L. & Aizenmann, E. *Science* **239**, 1293–1296 (1988).
20. Lohof, A. M., Quillan, M., Dan, Y. & Poo, M.-m. *J. Neurosci.* **12**, 1253–1261 (1992).
21. Xie, Z. P. & Poo, M.-m. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7069–7073 (1986).
22. Evers, J., Laser, M., Sun, Y., Xie, Z. P. & Poo, M.-m. *J. Neurosci.* **9**, 1523–1539 (1989).
23. Patel, N. & Poo, M.-m. *J. Neurosci.* **4**, 2939–2947 (1984).
24. Mattson, M. P. & Kater, S. B. *J. Neurosci.* **7**, 4034–4043 (1987).
25. Rehder, V. & Kater, S. B. *J. Neurosci.* **12**, 3175–3186 (1992).
26. Davenport, R. W. & Kater, S. B. *Neuron* **9**, 405–416 (1992).
27. Bedlack, R. S. Jr, Wei, M.-d. & Loew, L. M. *Neuron* **9**, 393–403 (1992).
28. Vernino, S., Amador, M., Luetje, C. W., Patrick, J. & Dani, J. A. *Neuron* **8**, 127–134 (1992).
29. Vijayaraghavan, S., Pugh, P. C., Zhang, Z. W., Rathouz, M. M. & Berg, D. K. *Neuron* **8**, 353–362 (1992).
30. Bixby, J. L. & Spitzer, N. C. *Dev. Biol.* **106**, 89–96 (1984).
31. Tokumitsu, H. et al. *J. Biol. Chem.* **265**, 4315–4320 (1990).
32. Ito, I., Hidaka, H. & Sugiyama, H. *Neurosci. Lett.* **121**, 119–121 (1991).
33. Gryniewicz, G., Poenie, M. & Tsien, R. Y. *J. Biol. Chem.* **260**, 3440–3450 (1985).
34. Connor, J. A. *Cell Calcium* **14**, 173–188 (1993).
35. Connor, J. A. & Nikolakopoulou, G. *Lec. Math. Life Sci.* **15**, 79–101 (1982).
36. Cohan, C. S., Connor, J. A. & Kater, S. B. *J. Neurosci.* **7**, 3588–3599 (1987).
37. McCobb, D. P., Cohan, C. S., Connor, J. A. & Kater, S. B. *Neuron* **1**, 377–385 (1988).
38. Spitzer, C. C. & Lamborghini, H. E. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1641–1645 (1976).
39. Anderson, M. J., Cohen, M. W. & Zorychta, E. J. *Physiol. Lond.* **268**, 731–758 (1977).
40. Tabti, N. & Poo, M.-m. in *Culturing Nerve Cells* (eds Banker, G. & Goslin, K.) 137–153 (MIT Press, Cambridge, 1991).

ACKNOWLEDGEMENTS. This work was supported by a grant from the NIH.

Changing subunit composition of heteromeric NMDA receptors during development of rat cortex

Morgan Sheng, Jennifer Cummings,
Leslie Ann Roldan, Yuh Nung Jan & Lily Yeh Jan

Howard Hughes Medical Institute and Departments of Physiology and Biochemistry, University of California, San Francisco, California 94143-0724, USA

ACTIVATION of the *N*-methyl-D-aspartate (NMDA) receptor is important for certain forms of activity-dependent synaptic plasticity, such as long-term potentiation (reviewed in ref. 1), and the patterning of connections during development of the visual system (reviewed in refs 2, 3). Several subunits of the NMDA receptor have been cloned: these are NMDAR1 (NR1), and NMDAR2A, 2B, 2C and 2D (NR2A–D)^{4–8}. Based on heterologous co-expression studies, it is inferred that NR1 encodes an essential subunit of NMDA receptors and that functional diversity of NMDA receptors *in vivo* is effected by differential incorporation of subunits NR2A–NR2D^{5–8}. Little is known, however, about the actual subunit composition or heterogeneity of NMDA receptors in the brain. By co-immunoprecipitation with subunit-specific antibodies, we present here direct evidence that NMDA receptors exist in rat neocortex as heteromeric complexes of considerable heterogeneity, some containing both NR2A and NR2B subunits. A progressive alteration in subunit composition seen postnatally could contribute to NMDA-receptor variation and changing synaptic plasticity during cortical development.

NR1-FP antibodies, which are raised against NR1 fusion proteins and should recognize all known splice forms of the NR1 subunit^{9,10}, bind to a heterogeneous band of relative molecular mass (M_r) ~120K on immunoblots of rat brain membranes (Fig. 1), in agreement with the reported size of the NR1 protein^{11,12}. NR1-C1 and NR1-N1 antibodies, raised against alternatively spliced exons of NR1, recognize distinct, though overlapping, subsets of the bands recognized by NR1-FP, conwith their specificity for different overlapping subsets of splice variants of NR1^{9,10}. NR2A and NR2B antibodies recognize bands of M_r ~165K and ~170K, respectively, consistent with predicted sizes of NR2A and NR2B^{5–8}.

After solubilization of cortical membranes with 'RIPA' buffer (see Fig. 2 legend), NR1-FP, NR1-C1 and NR1-N1 antibodies immunoprecipitate not only NR1 polypeptides, as expected, but also significant amounts of NR2A and NR2B (Fig. 2). Conversely, NR2A and NR2B antibodies immunoprecipitate NR1 proteins in addition to their respective cognate antigens. The mutually consistent results obtained with five independent anti-

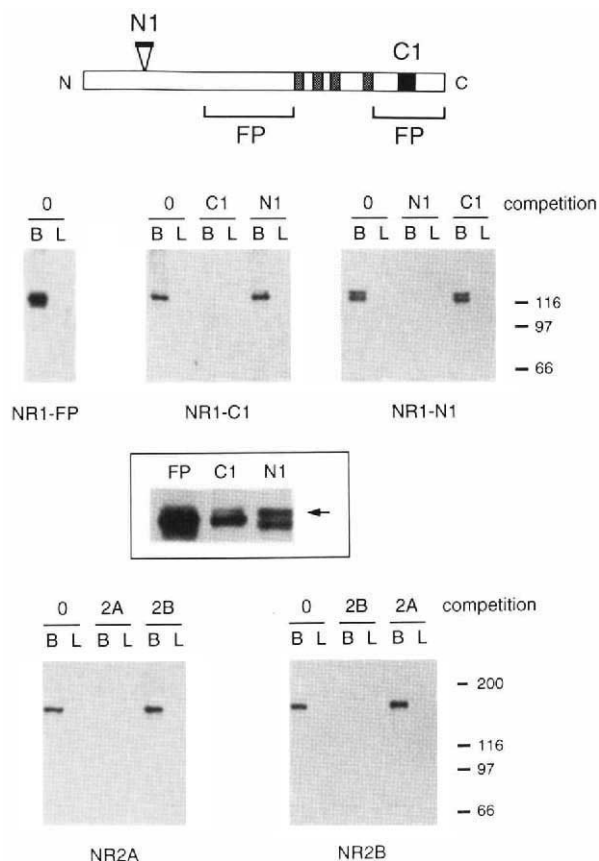


FIG. 1 NMDA-receptor subunit-specific antibodies. Top section: Diagram of the NR1 protein, showing positions of alternatively spliced N1 and C1 exons. Brackets indicate the segments of NR1 incorporated in fusion proteins for raising NR1-FP antibodies. Putative transmembrane segments are shown (stippled). Middle and bottom sections: Immunoblots of membrane proteins from whole rat brain (B) or rat liver (L) were probed with affinity-purified antibodies to NR1-FP, C1, N1, NR2A or NR2B, as indicated at the bottom of each panel. Immunoblots were probed in the absence of competitor (0), or in the presence of various competitor peptides (10 µg ml⁻¹), as indicated. Boxed inset shows in detail the bands recognized by NR1-FP, C1 and N1 antibodies. The doublets recognized by C1- and N1-exon-specific antibodies correspond to different but overlapping subsets of the heterogeneous band(s) recognized by the NR1-FP antibodies. The upper co-migrating band recognized by both C1 and N1 antibodies (arrow) probably contains NR1 polypeptides which incorporate both the C1 and N1 exons (see text). METHODS. Rabbit antibodies were raised against the following peptides: C1:CDKSGRAEPDPKPKATFRAITSTLASSFKRRRTSKDT (residues 864–900 of NMDAR1, ref. 4, corresponding to exon 5, ref. 9). N1:SKKRYNENLDQLSYDNKRGPKC, corresponding to exon 21, ref. 9. NR2A:CDNILDKPREIDLSRPSRSISLKDRELL (residues 1,293–1,321)⁸. NR2B:SPHWEHVDLTDIYKERSDDFKRDSVSGGGGPC (residues 1,143–1,173)⁸. Underlined cysteine residues were added for coupling to carrier. NR1-FP antibodies were raised against two bacterial fusion proteins containing residues 339–561 and 834–938 of NMDAR1 (ref. 4) inserted in-frame in vector pRSETB (Invitrogen), and induced and purified on zinc columns according to the manufacturer's directions. Kv1.4 antibodies have been described²⁶. All antibodies were affinity-purified using their respective immunogens coupled to Sulfolink columns (Pierce) or Affi-Gel 10 (Biorad). Immunoblots were performed as described²⁷.

bodies make it unlikely that co-immunoprecipitation of these NMDA-receptor subunits resulted from antibody crossreactivity. Control immunoprecipitations indicate that the potassium ion channel subunit Kv1.4 is not associated with NR1, NR2A or NR2B.

Under strongly denaturing conditions of receptor solubilization, the subunit-specific antibodies effectively immunoprecipitate their cognate antigens, but co-immunoprecipitation of other subunits is virtually abolished (Fig. 2). Together, these results indicate that NR1, NR2A and NR2B are associated through non-covalent interactions, and imply the existence of heteromeric NMDA receptors in rat brain. The demonstration of subunit association between NR1 and NR2A or NR2B subunits *in vivo* is consistent with the prediction that at least some native NMDA receptors consist of hetero-oligomers of NR1 with NR2A or NR2B¹³. It is interesting, however, that in addition to immunoprecipitating their respective cognate antigens and NR1, NR2A antibodies efficiently immunoprecipitate NR2B, and conversely, NR2B antibodies also immunoprecipitate NR2A (Fig. 2). This co-immunoprecipitation, which is prevented by prior denaturation of the membrane proteins, indicates that at least a portion of the NR2A and NR2B subunits are present in the same complex *in vivo*. Given that the NR2A/NR2B heteromeric combination cannot form functional NMDA receptors in the absence of NR1^{5,8}, the detection of native complexes containing both NR2A and NR2B suggests that a proportion of cortical NMDA receptors are hetero-oligomers containing at least three distinct subunits (NR1/NR2A/NR2B). The functional characteristics of 'triple subunit' heteromeric NMDA receptors incorporating both NR2A and NR2B subunits remain to be determined.

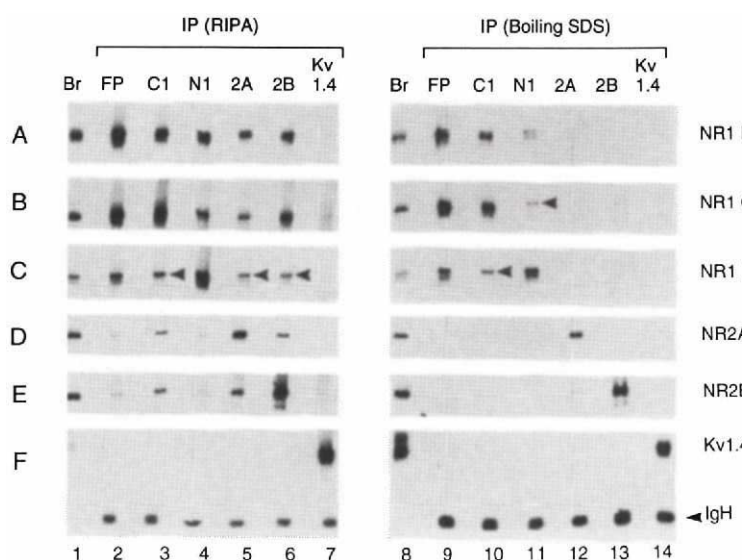
A change in functional properties of the NMDA receptor has been invoked as a potential mechanism contributing to the loss of synaptic plasticity during brain maturation¹⁴⁻¹⁹. RNA and immunoblot analysis demonstrates that NMDA receptor subunits are differentially expressed during postnatal cortical development (Fig. 3). Expression of NR1 mRNAs and proteins (total NR1, N1-exon-containing NR1 (N1⁺-NR1) and C1-exon-containing NR1 (C1⁺-NR1)) follow similar temporal profiles: levels are relatively low at birth, rise to a peak between two and three

weeks, and subsequently fall to an intermediate plateau with further ageing. The NR2B messenger RNA and polypeptide, however, are already highly expressed within a day of birth (P1) and remain roughly constant into adulthood. In contrast, NR2A mRNA and protein are undetectable at birth, and increase progressively over the next three weeks to reach its adult plateau. These results correlate with *in situ* hybridization studies of NMDA-receptor expression during development²⁰.

To test directly the idea that changing relative levels of NR1, NR2A and NR2B subunits result in a developmental change in the molecular composition of native NMDA receptors, co-immunoprecipitation assays were performed at different postnatal ages (Fig. 4). C1- and N1-specific antibodies co-immunoprecipitated NR2B from cortical membranes throughout the period tested (P1-P53), although the amount of NR2B co-immunoprecipitated was consistently much smaller for N1 than for C1 antibodies. The amount of NR2A co-immunoprecipitated with either C1 or N1 antibodies, however, was undetectable at P1, and progressively increased to adult levels by about three weeks of age (Fig. 4). A substantial fraction of newly expressed NR2A subunits seems to be incorporated into a receptor subtype containing all three subunits (NR1/NR2A/NR2B), as NR2A and NR2B are co-immunoprecipitated by antibodies to either subunit from P7 to adulthood (Fig. 2, and data not shown). However, the differential co-immunoprecipitation of NR2A and NR2B by N1 and C1 antibodies (discussed below) suggests that there must also be some segregation of NR2A and NR2B into three different NMDA-receptor subtypes.

During postnatal development, native NMDA receptors increasingly express low-affinity ifenprodil binding sites²¹, show a progressive shortening in the duration of NMDA-mediated excitatory postsynaptic currents^{14,15,17}; become less sensitive to glycine¹⁸, and become more susceptible to Mg²⁺ block^{16,18,22}. In this regard, it is probably significant that NR1/NR2A heteromers show a much lower affinity for ifenprodil than NR1/NR2B²¹, that agonist-evoked currents through the NR1/NR2A heteromer exhibit a faster decay rate than other heteromeric combinations⁵ (H. Monyer and B. Sakmann, personal communication), NR1/NR2A has lower affinity for glycine than does NR1/NR2B^{7,13}, and that protein kinase C potentiation of

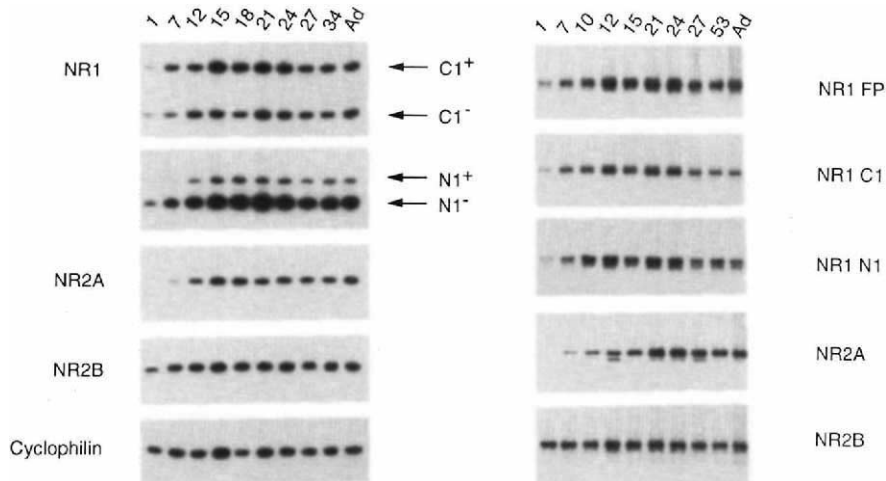
FIG. 2 Co-immunoprecipitation of NMDA-receptor subunits. Adult (P53) rat cortical membrane proteins solubilized in RIPA buffer (left panel), or in boiling SDS followed by dilution into RIPA (right panel), were immunoprecipitated by different affinity-purified antibodies, as indicated at the top of each column. The immunoprecipitates (IP) were separated by SDS 6% PAGE and transferred to nitrocellulose; the same nitrocellulose filter was then analysed by sequential immunoblotting using successive antibodies, as indicated at the right of each row. Lanes 1 and 8 ('Br') were loaded with total cortex membrane protein without immunoprecipitation. Bands which probably correspond to N1⁺C1⁺ splice forms of NR1 are indicated by arrows (see text). The signals derived from the heavy chain of the immunoprecipitating IgG (IgH) are shown for the Kv1.4 immunoblot only. Autoradiographic exposure times for rows A-E are identical for both panels; the right-hand panel of row F (Kv1.4 immunoblot) is exposed for three times longer than the left-hand panel. Because the results shown in each of lanes 2-7 and 9-14 are from the same single immunoprecipitation, the ratios of immunoreactivities detected by different immunoblotting antibodies can be compared between different columns as a measure of relative co-immunoprecipitation. **METHODS.** P53 rat cortical membranes, prepared as described²⁶, were solubilized in 1 × RIPA buffer (50 mM Tris, pH 7.6, 150 mM NaCl, 1 mM EDTA, 1% NP40, 0.5% deoxycholate, 0.1% SDS), or by boiling in 0.5% SDS (5 min) followed by addition of 1.25 × RIPA lacking SDS to obtain identical final RIPA buffer conditions. Insoluble material was pelleted at 110,000g for 60 min and equal amounts of supernatant were immunoprecipitated using affinity-purified antibodies



at 10 µg ml⁻¹, as described²⁷. The immunoprecipitates were analysed on immunoblots by repeatedly stripping the membrane and reprobing with successive antibodies²⁷ at a concentration of 1-2 µg ml⁻¹. Only half as much starting membrane protein (200 µg) was used for immunoprecipitation from boiling SDS extracts, as this method solubilized NMDA receptors better than RIPA alone.

FIG. 3 Differential expression of NMDA-receptor subunits during postnatal cortical development. Left panel, analysis of rat cortex RNA by quantitative reverse transcriptase-polymerase chain reaction (RT-PCR) assay, using oligonucleotide primers specific for NR1, NR2A or NR2B transcripts, as indicated. For NR1, amplification with primers flanking the C1 or N1 exon resulted in two different-sized products, corresponding to mRNAs containing, or lacking, these exons, as indicated. Amplification with cyclophilin-specific primers was used as an internal control for RNA/complementary DNA input. Right panel, Immunoblot analysis of rat cortex membrane proteins (~15 µg per lane), using subunit-specific antibodies, as indicated. The same immunoblot filter was successively probed, stripped and reprobed with different antibodies, as described for Fig. 2. Numbers indicate postnatal age of rats in days; the 'adult' rat was ~12 weeks old.

METHODS. Quantitative RT-PCR assay was performed as described in detail elsewhere²⁸. Identical (5%) aliquots of cDNA, reverse-transcribed from 2 µg DNase-pretreated cortical RNA of various ages, were amplified by PCR for 28 cycles with NMDA-receptor subunit primers, or for 22 cycles with cyclophilin primers, in the presence of ³²P-dCTP. Number of cycles was within the exponential (quantitative) range of amplification, as determined by preliminary experiments. Gene-specific primers, sense/antisense (and sizes of PCR amplification products) were:



NR1(flanking C1 exon): CAGATGCAGCTGGCTTTTGAG/CGGCAGCACTGTGTCTTTTGG (C1⁺, 214 base pairs (bp), C1⁻, 103 bp); NR1(flanking N1 exon): GGTTTGAGATGATGCGAGTCTAC/CAGCAGAGCCGTCACATTCTTGG (N1⁺, 235 bp; N1⁻, 172 bp); NR2A: CAATCTGACTGGATCAGAGC/CTGTCCCTTCCCTTGAAGGATC (207 bp); NR2B: CAATAACCCACCTGTGAGG/GGTGGGTTGTCACATCATAG (183 bp); cyclophilin: GACCAACACAAATGGTTC/CAGAAGGAATGGTTTGATGG (228 bp).

NMDA-receptor activity, thought to be mediated by reduced Mg²⁺ block, is only transient for NR1/NR2A but is sustained for NR1/NR2B⁷. These correlations suggest that the increasing incorporation of NR2A subunits could underlie some of the changes in NMDA-receptor properties observed during development. Such a change in the molecular composition of NMDA receptors could contribute to the loss of cortical plasticity that accompanies maturation of the brain.

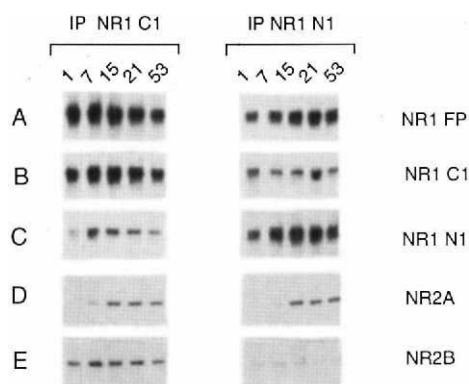


FIG. 4 Changing patterns of NMDA-receptor subunit co-immunoprecipitation during cortical development. Cortical membranes from various ages were solubilized in RIPA buffer and immunoprecipitations performed using NR1-C1 or -N1 antibodies, as indicated. Sequential immunoblotting was performed as for Fig. 2. Numbers refer to postnatal age in days. The amount of starting membrane protein for immunoprecipitation was increased at P1 (~200%) and P7 (~150%) relative to P15, P21 and P53 (200 µg total membrane protein), to compensate for the lower expression of NR1 at the earlier ages (see Fig. 3, right); thus, comparable amounts of NR1 (as measured by NR1-FP immunoreactivity) are immunoprecipitated at all time points. There was no significant co-immunoprecipitation of NR2A and NR2B after co-solubilization of P1 brain membranes (expressing NR1/NR2B) with adult cerebellar membranes (expressing NR1/NR2A), indicating that dissociation and reassociation of subunits did not occur artefactually under our solubilization and immunoprecipitation conditions. Methods were as for Fig. 2.

The co-immunoprecipitation data also provide evidence for differential association between the various NMDA-receptor subunits. For instance, although two major N1-exon-containing NR1 bands are detected on immunoblot, only the upper of these two N1⁺-NR1 bands is predominantly co-immunoprecipitated with NR2A and NR2B antibodies (Fig. 2, row C, arrows). The N1⁺-NR1 band that is co-immunoprecipitated with NR2A and NR2B almost certainly contains NR1 polypeptide(s) which incorporate both the N1 and C1 exons (that is, N1⁺C1⁺ - NR1), because, first, a single band of the same size is also detected in anti-C1 immunoprecipitates probed with N1-specific antibodies (Fig. 2, coordinate C3), and this band persists even after receptor denaturation (Fig. 2, coordinate C10), and, second, N1 antibodies immunoprecipitate a single band of identical size in denaturing conditions which is immunoreactive to C1-exon-specific antibodies (Fig. 2, coordinate B11). Thus, NR2A and NR2B are preferentially complexed with N1⁺C1⁺, rather than with N1⁺C1⁻ variants of NR1. A corollary of this is that N1⁺C1⁺ splice forms of NR1 are segregated to some degree from N1⁺C1⁻ variants, a conclusion supported by the finding that, even in non-denaturing conditions, C1 antibodies immunoprecipitate only the N1⁺C1⁺ upper band of the N1 doublet (Fig. 2, coordinate C3).

As a further example of subunit-specific associations, C1 antibodies consistently co-immunoprecipitate more NR2B protein than do N1 antibodies (Fig. 4, row E), even though both antibodies immunoprecipitate comparable amounts of total NR1 protein as assayed by NR1-FP immunoblotting (Fig. 4, row A). In contrast, N1 and C1 antibodies are similarly efficient for co-immunoprecipitation of NR2A (Fig. 4, row D). Because splice variants of NR1 have distinct functional properties^{9-10,23,25}, differential association of specific NR1 splice forms with NR2A and NR2B could be a means of generating further diversity of NMDA receptor subtypes *in vivo*. □

Received 17 September; accepted 23 December 1993.

1. Bliss, F. V. P. & Collingridge, G. L. *Nature* **361**, 31-39 (1993).
2. Constantine-Paton, M., Cline, H. T. & Debski, E. A. *Rev. Neurosci.* **13**, 129-154 (1990).
3. Schatz, C. J. *Neuron* **5**, 745-756 (1990).
4. Moriyoshi, K. et al. *Nature* **354**, 31-37 (1991).

5. Monyer, H. et al. *Science* **256**, 1217–1221 (1992).
6. Meguro, H. et al. *Nature* **357**, 70–74 (1992).
7. Kutsuwada, T. et al. *Nature* **358**, 36–41 (1992).
8. Ishii, T. et al. *J. Biol. Chem.* **268**, 2836–2843 (1993).
9. Hollmann, M. et al. *Neuron* **10**, 943–954 (1993).
10. Sugihara, H. et al. *Biochem. biophys. Res. Commun.* **185**, 826–832 (1992).
11. Tingley, W. G., Roche, K. W., Thompson, A. K. & Huganir, R. L. *Nature* **364**, 70–73 (1993).
12. Chazot, P. L., Cik, M. & Stephenson, F. A. *J. Neurochem.* **59**, 1176–1178 (1992).
13. Stern, P., Behe, P., Schoepfer, R. & Colquhoun, D. *Proc. R. Soc. Lond. B* **250**, 271–277 (1992).
14. Carmignoto, G. & Vicini, S. *Science* **258**, 1007–1011 (1992).
15. Hestrin, S. *Nature* **357**, 686–689 (1992).
16. Ben-Ari, Y. et al. *Neurosci. Lett.* **94**, 88–92 (1988).
17. Kato, N., Artola, A. & Singer, W. *Dev. Brain Res.* **60**, 43–50 (1991).
18. Kleckner, N. W. & Dingledine, R. *Molec. Brain Res.* **11**, 151–159 (1991).
19. Tsumoto, T., Hagihara, K., Sato, H. & Hata, Y. *Nature* **327**, 513–514 (1987).

20. Watanabe, M. et al. *Neuroreport* **3**, 1138–1140 (1992).
21. Williams, K., Russell, S. L., Shen, Y. M. & Molinoff, P. B. *Neuron* **10**, 267–278 (1993).
22. Kato, N. & Yoshimura, H. *Proc. natn. Acad. Sci. U.S.A.* **90**, 7114–7118 (1993).
23. Nakanishi, N., Axel, R. & Schneider, R. *Proc. natn. Acad. Sci. U.S.A.* **89**, 8552–8556 (1992).
24. Durand, G. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 9359–9363 (1992).
25. Anantharam, V. et al. *FEBS Lett.* **305**, 27–30 (1992).
26. Sheng, M., Tsaur, M.-L., Jan, Y. N. & Jan, L. Y. *Neuron* **9**, 271–284 (1992).
27. Sheng, M., Liao, Y. J., Jan, Y. N. & Jan, L. Y. *Nature* **365**, 72–75 (1993).
28. Tsaur, M.-L., et al. *Neuron* **8**, 1055–1067 (1992).

ACKNOWLEDGEMENTS. We thank T. J. Baldwin, M. Li, G. Feger and M. Hattori for help with antibody preparation, and B. Zhao, C.-L. Huang and M. Stryker for critical reading of the manuscript. This work was supported by the Silvio Conte Center grant from the National Institute for Mental Health, Y.N.J. and L.Y.J. are Howard Hughes Investigators, and M.S. is a recipient of a HHMI Postdoctoral Fellowship for Physicians. J.C. is supported by a HHMI predoctoral fellowship.

Mice lacking brain-derived neurotrophic factor develop with sensory deficits

Patrik Ernfors, Kuo-Fen Lee & Rudolf Jaenisch

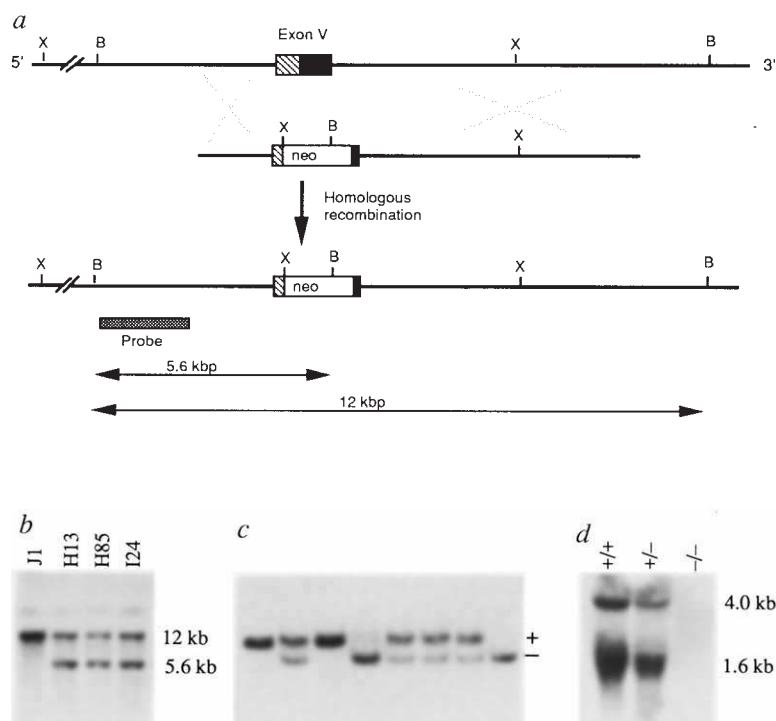
Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA

DURING vertebrate development, neuronal survival depends on target-derived neurotrophic factors^{1,2}. Brain-derived neurotrophic factor³ (BDNF), a member of the neurotrophin family, can prevent the death of particular peripheral sensory neurons *in vitro*^{4–6}, and of central motor neurons as well as dopaminergic and cholinergic

neurons of the basal forebrain during development^{7–9}. It also prevents the death of motor neurons and midbrain dopaminergic neurons induced by lesions^{8,10–12}. Here we show that mutant mice lacking BDNF have severe deficiencies in coordination and balance, associated with excessive degeneration in several sensory ganglia including the vestibular ganglion. The few remaining vestibular axons fail to contact the vestibular sensory epithelia, and terminate in the adjacent connective tissue. Survival of sympathetic, midbrain dopaminergic and motor neurons is not affected. These results indicate that BDNF is required for the survival and target innervation of particular neuronal populations.

A targeting vector containing a deleted gene encoding BDNF (Fig. 1a) was used to transfect embryonic stem (ES) cells (Fig. 1b) for establishing mutant mice (Fig. 1c). Messenger RNA isolated from cerebral cortex, where BDNF mRNA is abundantly expressed, was hybridized to a BDNF complementary DNA fragment. Mice heterozygous for the mutation showed a reduc-

FIG. 1 Targeted inactivation of the BDNF gene in mouse ES cells and generation of mutant mice. **a**, Schematic diagram of the strategy used to target the BDNF gene. Top line, genomic structure of the BDNF gene; the entire coding sequences are contained within exon 5 represented by a box, prepro-coding sequences are striped, and the black box represents coding sequences for the processed mature biologically active protein. The 9.5-kb targeting replacement vector used for electroporation is shown in the middle. A neo expression cassette, including the phosphoglycerate kinase promoter and polyadenylation signals, was inserted into a partially deleted exon 5 of the BDNF gene. The deletion left 49 and 30 amino acids in the N and C termini, respectively. The embryonic stem cell line J1 was electroporated with the targeting vector followed by selection with G418 (0.4 mg ml⁻¹). G418-resistant clones were expanded and analysed by Southern blot analysis. Bottom line represents the structure of the inactivated BDNF gene. X, XbaI; B, BglII. **b**, Southern blot of three ES clones. DNA from parental J1 ES cells and mutant clones H13, H85 and I24 was digested with BglII and hybridized to a genomic probe external to the 5' homologous arm. This probe hybridized to a 12-kb and a 5.6-kb fragment from wild-type and mutant alleles, respectively. Clones that contained the expected 5.6-kb BglII fragment diagnostic for homologous recombinants were obtained at a frequency of 1 in 24. These ES clones were further analysed by digestion with XbaI, generating a 13-kb wild-type and a 7.2-kb mutant fragment when probed. The membranes were rehybridized to a neo probe to verify a single integration. Of seven homologously recombined ES clones injected into blastocysts, six contributed to the germ line of recipient embryos. **c**, Southern blot of the offspring from a BDNF^{+/+} × BDNF^{+/+} cross. Tail DNA was extracted and digested with BglII and analysed as described in **b**. **d**, Northern blot of BDNF mRNA expression in the cerebral cortex of BDNF^{+/+}, BDNF^{+/−} and BDNF^{−/−} mice. Total RNA (20 µg) were analysed from each sample. Equal loading was verified by ethidium bromide staining of the gel and by reprobing of the membrane with a β-tubulin cDNA.



METHODS. The J1 ES cell line was derived and grown on embryonic fibroblast feeder cells as described²². Following electroporation of cells with 60 µg ml⁻¹ of linearized targeting vector (Bio-Rad gene pulser at settings of 400V, 25 µF) cells were selected in 0.4 mg ml⁻¹ G418. Resistant colonies were trypsinized individually. Half of the cells was expanded for freezing and the other half for genotyping. Mice were generated as described²².

tion of BDNF mRNA to about half the levels from wild-type mice (Fig. 1*d*), but displayed no overt behavioural abnormality. Homozygous mutant mice lacked detectable levels of BDNF mRNA (Fig. 1*d*) and were reduced in size as compared to wild-type littermates; most died during the second postnatal week. They showed defective coordination of movements and balance, with head bobbing and tilting, and spinning during periods of hyperactivity, followed by immobility; they only rarely survived to adulthood. The same phenotype was seen in inbred 129 ter/Sv mice and in crosses to BALB/c.

At autopsy, several peripheral sensory ganglia were markedly smaller in mutant mice than in controls. To quantify neuronal loss, neurons were counted in serial sections of the dorsal root, trigeminal, vestibular, nodose and superior cervical ganglia, mesencephalic trigeminal nucleus and the facial motor nucleus (Table 1). The number of neurons was significantly reduced, with $70 \pm 3\%$ remaining in the dorsal root ganglion, $56 \pm 2\%$ in the trigeminal ganglion, $52 \pm 2\%$ in the mesencephalic trigeminal nucleus, $18 \pm 4\%$ in the vestibular ganglion and $34 \pm 3\%$ in the nodose ganglia. The numbers of facial motor neurons and sympathetic superior cervical ganglion neurons were not significantly different from those of wild-type mice ($97 \pm 6\%$ and $114 \pm 6\%$, respectively). Whereas the morphology of neurons in the ganglia

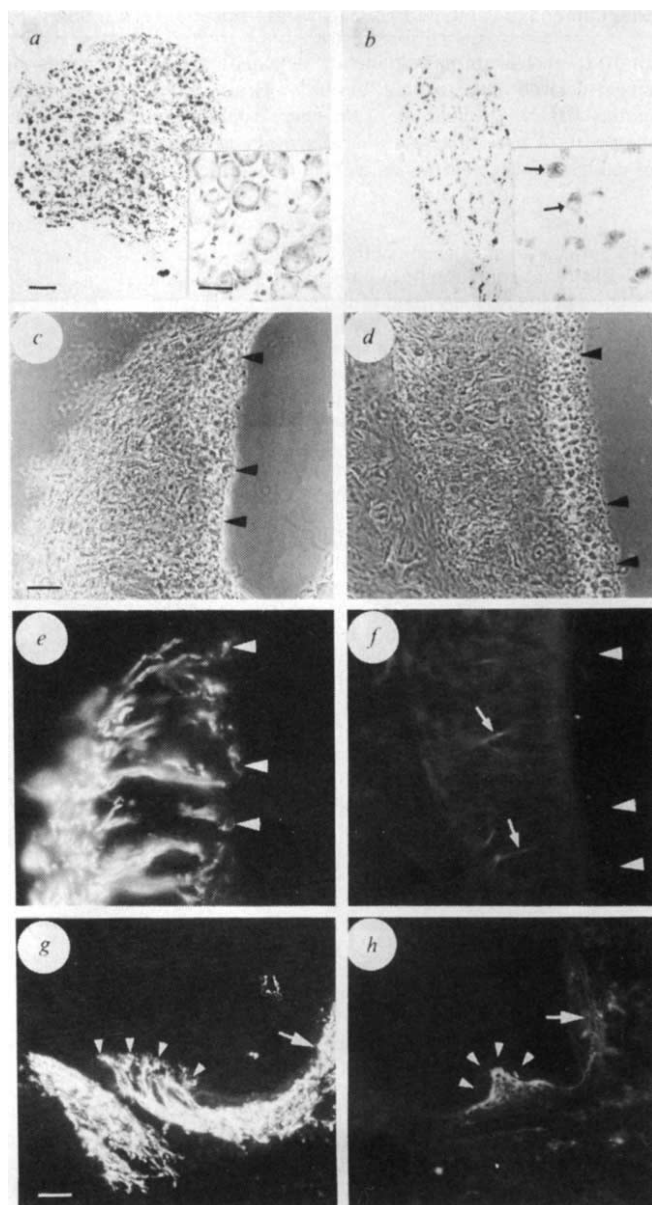
of mutant mice, including spiral ganglion neurons, was normal, vestibular ganglion neurons appeared atrophic, with small cell size, and often contained a displaced nucleus (Fig. 2*a* and *b*, and insets). Thus, lack of BDNF leads to excessive cell death in all populations of primary sensory neurons examined.

To determine whether the loss of neurons in the postnatal mutant mice was caused by a failure of ganglion formation or extension of nerve processes before programmed cell death, the nervous system of whole embryos was stained with a monoclonal antibody raised against neurofilament (M_r 160K). No differences were detected in sizes of, or in the central and peripheral nerve processes from, the trigeminal, geniculate, petrose and nodose ganglia at embryonic days 10.5 and 11.5 in wild-type (Fig. 3*a*) or mutant embryos (Fig. 3*b*). Thus, lack of BDNF did not affect ganglion formation or extension and target encounter of central and peripheral nerve processes during early developmental stages.

To study whether circling and spinning and the defects in coordination of movements could be associated with abnormal labyrinth innervation, an antibody (RT97) directed against phosphorylated epitopes on neurofilaments was used to investigate innervation of the inner ear. In the inner ear of wild-type mice, dense RT97 immunoreactive fibre bundles entered the ves-

FIG. 2 Lack of BDNF leads to excessive cell death and atrophy of vestibular ganglion neurons as well as deficiency of inner ear innervation. *a*, *c*, *e* and *g*, Wild-type; *b*, *d*, *f* and *h*, BDNF(-/-) mice. *a* and *b*, Cresyl-violet-stained sections of the vestibular ganglion. Insets, high magnification of vestibular ganglion neurons. Arrows in *b*, indicate small neurons with displaced nuclei. *c* and *d*, Phase-contrast micrographs of the saccular macula; arrowheads indicate the sensory epithelia. *e* and *f*, Same sections as in *c* and *d*, with monoclonal antibody RT97 directed against phosphorylated epitopes on neurofilaments. Arrowheads in *e* indicate nerve chalice in the sensory epithelia, and in *f* the sensory epithelia; arrows in *f* indicate fibres terminating in the connective tissue. *g* and *h*, Immunohistochemistry with RT97 in the crista of the semicircular duct. Arrow indicates the nerve trunk projecting to the crista. Arrowheads indicate the sensory epithelia. Note the overall low immunoreactivity in *f* and *h* as compared to *e* and *g*, and that nerve fibres terminate in the connective tissue in *f* and *h*. Scale bars: in *a* and *b*, 70 μ m; in insets, 20 μ m; in *c*-*f*, 20 μ m; in *g* and *h*, 40 μ m.

METHODS. Postnatal day 14 BDNF(-/-) and wild-type animals were perfused with 4% paraformaldehyde. Vestibular ganglia and inner ears were dissected and postfixed for 2 h. Vestibular ganglia were equilibrated with 30% sucrose, sectioned at 7 μ m and stained with cresyl violet. Inner ears were decalcified, equilibrated with sucrose and sectioned at 10 μ m. Sections were preincubated in dilution buffer (0.5 M NaCl, 0.01 M phosphate buffer, pH 7.3, 3% bovine serum albumin and 0.3% Triton X-100) for 1 h, followed by overnight incubation with mouse RT97 monoclonal antibody²³ diluted at 1:500 in dilution buffer containing 5% goat serum. After 4 washes in PBS, sections were incubated for 2 h with a rhodamine-conjugated goat anti-mouse secondary antiserum (1:300 dilution).



tibule and the ampulla of the semicircular ducts (Fig. 2e, g), transversed through the connective tissue underlying the sensory epithelia and innervated the hair cells of the sensory epithelia (Fig. 2e and g, arrowheads). Much less RT97 immunoreactivity was detected in mutant mice. Only small fibre bundles were entering the sacculus (Fig. 2f), utricle and the ampulla of the semicircular ducts (Fig. 2h, arrow). Furthermore, most, if not all, of the remaining fibres terminated in the connective tissue adjacent to the sensory epithelia of the sacculus (Fig. 2f, arrowheads), utricle and of the crista of the semicircular ducts (Fig. 2h). On no occasion were nerve chalice found, characteristic of innervation of type I hair cells. The structural integrity of the organ of corti, macula of the sacculus (Fig. 2c and d) and the utricle, and the crista of the semicircular ducts appeared normal in mutant mice. Innervation of the cochlear inner and outer hair cells was normal, and the mice responded to auditory stimuli. Thus at postnatal day 14, lack of BDNF led to marked cell death in the vestibular ganglion and the remaining axons did not contact the target cells, which express BDNF and neurotrophin-3 (NT3) mRNAs^{13,14}. This suggests that BDNF acts at the cellular level and is necessary for establishing or maintaining innervation in the vestibular compartments of the inner ear. We cannot exclude the possibility that phosphorylation of neurofilament protein epitopes recognized by RT97 is downregulated in the mutant mice. However, the few afferents exiting the mutant ganglia were as intensely RT97-immunoreactive as those of wild-type ganglia. This suggests that survival rather than phosphorylation of the large neurofilament protein is dependent on BDNF. The behavioural abnormalities, including circling and spinning are an expected consequence of defective vestibular innervation and are typically seen in mutant mice such as

waltzer, shaker or jerker, which have innervation defects or malformation of the inner ear¹⁵.

Because BDNF promotes the survival of embryonic central dopaminergic neurons in culture⁸, this population of neurons was examined in mutant mice by immunocytochemistry using an antiserum against tyrosine hydroxylase. There was no apparent loss of central dopaminergic neurons in BDNF mutant mice (Fig. 3d) when compared to wild-type mice (Fig. 3c). In addition, cresyl violet stain of sections from mutant brains revealed no gross abnormalities in the hippocampus or the cerebellum, neurons of which respond to BDNF in culture^{16,17}.

Our results show that lack of BDNF expression does not affect the initial formation of cranial ganglia, extension of nerve processes and target encounter, but leads at later developmental stages to excessive cell death of particular populations of peripheral neurons, and a failure to establish or to maintain target innervation. Lack of BDNF results in loss of neurons in all sensory ganglia examined. These *in vivo* results are consistent with previous studies showing that BDNF promotes the survival of cultured peripheral sensory neurons⁴⁻⁷ and enhances the survival of dorsal root and nodose ganglia neurons following infusion of BDNF into embryos^{18,19}. In contrast to previous studies showing survival effects of BDNF on embryonic central dopaminergic neurons in culture⁸, and on motor neurons following exogenous administration at the time of programmed cell death in the chick or following nerve transection in the postnatal rat⁹⁻¹², no obvious defects were seen in these populations of neurons in the BDNF mutant mice. These results suggest that *in vitro* results or manipulations in prenatal or postnatal animals may not always reflect the physiological role of the protein *in vivo*, or that the function of BDNF is compensated by other neurotrophins in

FIG. 3 Lack of BDNF does not affect peripheral ganglia formation or central dopaminergic neurons. Whole-mount immunohistochemistry for neurofilament in a, wild-type, and b, BDNF^{-/-} E11.5 embryos. No apparent abnormality of cranial ganglia and nerve processes was observed in the mutant embryo. Tyrosine hydroxylase immunohistochemistry of sections through the midbrain of c, wild-type, and d, BDNF^{-/-} 2-week-old mice. No apparent abnormality was observed in the mutant mouse. Abbreviations: g, geniculate ganglion; n, nodose ganglion; p, petrosal ganglion; snr, substantia nigra reticulata; snc, substantia nigra compacta; t, trigeminal ganglion; VTA, ventral tegmental area; Scale bar in c and d, 40 μ m.

METHODS. For whole-mount immunocytochemistry, E11.5 embryos were dissected, fixed for 2 h in 4% paraformaldehyde, washed three times (1 h each) in washing solution (1% Triton X-100, 1% goat serum in PBS) and incubated with a mouse anti-160K neurofilament monoclonal antibody (Sigma), diluted 1/300 in washing solution. After 6 d, embryos were washed and incubated overnight with a horseradish peroxidase-conjugated goat anti-mouse secondary antibody, washed, incubated with 0.5 mg ml⁻¹ diaminobenzidine in 50 mM Tris-HCl, pH 7.4, for 3 h in the dark, followed by a 6-min incubation with diaminobenzidine containing 0.010% hydrogen peroxidase. For tyrosine hydroxylase staining, mid-brain sections of P14 mice were processed as described in Fig. 2 legend and stained with a rabbit anti-tyrosine hydroxylase antibody (Peel-Freeze) (1:200) followed by a rhodamine-conjugated goat anti-rabbit secondary antibody.

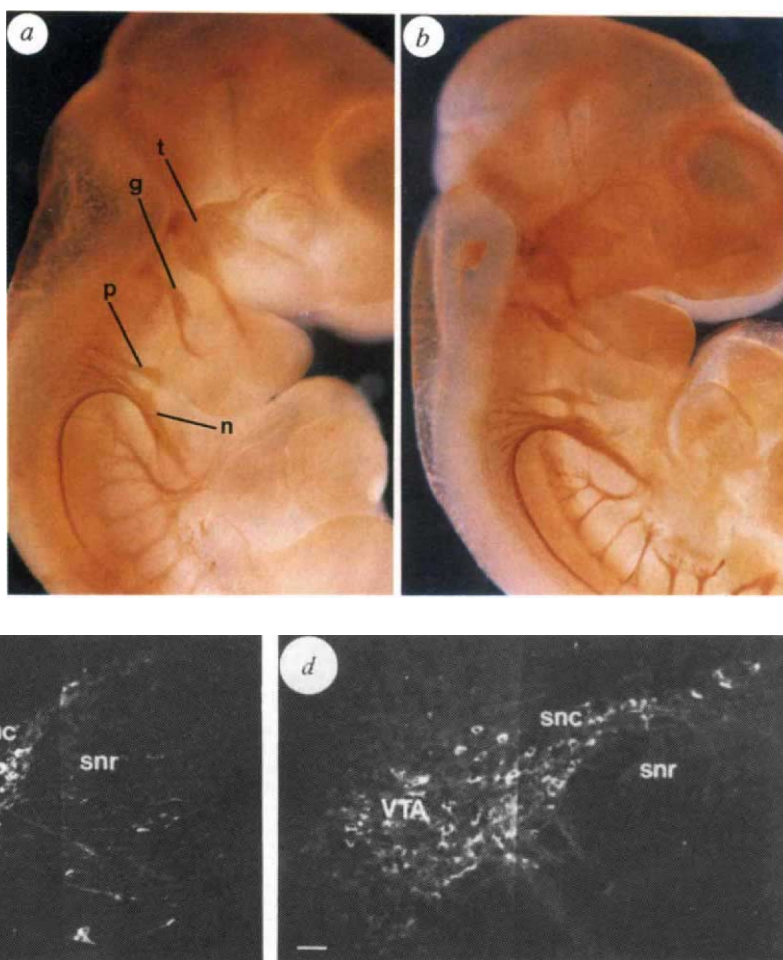


TABLE 1 Neuronal cell counts in sensory and sympathetic ganglia and facial motor nuclei of wild-type and BDNF mutant mice

	Mean no. of neurons $\pm$ s.e.m.	Per cent of control $\pm$ s.e.m.	
Trigeminal ganglion			
Wild type	29,221 $\pm$ 2,277	100 $\pm$ 8	(n=3)
BDNF -/-	16,505 $\pm$ 627	56 $\pm$ 2***	(n=3)
Me5			
Wild type	1,019 $\pm$ 31	100 $\pm$ 3	(n=6)
BDNF -/-	520 $\pm$ 25	52 $\pm$ 2***	(n=5)
Vestibular ganglion			
Wild type	2,301 $\pm$ 76	100 $\pm$ 3	(n=3)
BDNF -/-	75 $\pm$ 11	18 $\pm$ 4***	(n=3)
Nodose ganglion			
Wild type	3,931 $\pm$ 223	100 $\pm$ 5	(n=3)
BDNF -/-	1,325 $\pm$ 82	34 $\pm$ 3***	(n=4)
L4 DRG			
Wild type	4,402 $\pm$ 166	100 $\pm$ 4	(n=3)
BDNF -/-	3,104 $\pm$ 124	70 $\pm$ 3*	(n=4)
SCG			
Wild type	13,804 $\pm$ 1,302	100 $\pm$ 9	(n=3)
BDNF -/-	15,806 $\pm$ 798	114 $\pm$ 6	(n=3)
Facial motor neurons			
Wild type	2,213 $\pm$ 178	100 $\pm$ 8	(n=3)
BDNF -/-	2,146 $\pm$ 128	97 $\pm$ 6	(n=3)

Mice (postnatal days 14–16) were perfused with 4% paraformaldehyde. The trigeminal, vestibular, dorsal root, nodose and superior cervical ganglion and the brain were dissected, postfixed for 2–24 h, equilibrated with 30% sucrose and sectioned at 7 μ m thickness. Neurons with a clear nucleus and nucleoli were counted in every sixth section. Values were not corrected for split nucleoli. * $P < 0.05$, *** $P < 0.001$, Student's *t*-test. Abbreviations: L4 DRG, level L4 dorsal root ganglion; Me5, mesencephalic trigeminal nucleus; SCG, superior cervical ganglion.

the mutant mice. For example, BDNF, NT3 and neurotrophin-4 (NT4) all promote motor neuron survival in culture²⁰. Targeted inactivation of the common BDNF and NT4 receptor *trkB* leads to motor neuron degeneration²¹, suggesting that perhaps NT4 supports motor neuron survival *in vivo*. The availability of a BDNF-deficient mouse may not only further our understanding of the role of BDNF during development of the nervous system, but may also be useful to evaluate BDNF's potential as a treatment for neurological diseases. □

Received 9 December 1993; accepted 4 February 1994.

- Levi-Montalcini, R. *Science* **237**, 1154–1162 (1987).
- Barde, Y.-A. *Neuron* **2**, 1525–1534 (1989).
- Barde, Y.-A., Edgar, D. & Thoenen, H. *EMBO J.* **1**, 549–553 (1982).
- Lindsay, R. M., Thoenen, H. & Barde, Y.-A. *Dev Biol.* **112**, 319–328 (1985).
- Davies, A. M., Thoenen, H. & Barde, Y.-A. *J. Neurosci.* **6**, 1897–1904 (1986).
- Avila, M. A. et al. *Dev Biol.* **159**, 266–275 (1993).
- Alderson, R. F., Alterman, A. L., Barde, Y.-A. & Lindsay, R. M. *Neuron* **5**, 297–306 (1990).
- Hyman, C. et al. *Nature* **350**, 230–232 (1991).
- Oppenheim, R. W., Qin-Wei, Y., Prevett, D. & Yan, Q. *Nature* **360**, 755–757 (1992).
- Sendtner, M., Holtmann, B., Kolbeck, R., Thoenen, H. & Barde, Y. *Nature* **360**, 757–759 (1992).
- Yan, Q., Elliot, J. & Snider, W. D. *Nature* **360**, 753–755 (1992).
- Koliatsos, V. E., Clatterbuck, R. E., Winslow, J. W., Cayouette, M. H. & Price, D. L. *Neuron* **10**, 359–367 (1993).
- Ernfors, P., Merilo, J.-P. & Persson, H. *Eur. J. Neurosci.* **4**, 1140–1158 (1992).
- Pirvola, U. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 9915–9919 (1992).
- Lyon, M. F. & Searle, A. G. (eds) *Genetic Variant and Strains of the Laboratory Mouse* 2nd edn (Oxford Univ. Press, Oxford, 1989).
- Ip, N. Y., Li, Y., Yancopoulos, G. D. & Lindsay, R. M. *J. Neurosci.* **13**, 3394–3405 (1993).
- Segal, R. A., Takahashi, H. & McKay, R. D. *Neuron* **9**, 1041–1052 (1992).
- Hofer, M. M. & Barde, Y.-A. *Nature* **331**, 261–262 (1988).
- Kalchauer, C., Barde, Y.-A., Thoenen, H. & LeDouran, N. M. *EMBO J.* **6**, 2871–2873 (1987).
- Henderson, C. E. et al. *Nature* **363**, 266–270 (1993).
- Klein, R. et al. *Cell* **75**, 113–122 (1993).
- Li, E., Bestor, T. H. & Jaenisch, R. *Cell* **69**, 915–926 (1992).
- Wood, J. N. & Anderton, B. H. *Biosci. Rep.* **1**, 263–268 (1981).

ACKNOWLEDGEMENTS. We thank J. Daussman, R. Currey for technical assistance; G. Kemske for secretarial assistance; P. Wistrand and H. Aldskogius for the gift of RT97 antibody; and H. Gardner and B. Bates for discussion and for reading the manuscript. This research was supported by the NIH. P.E. was supported by a fellowship from the Swedish MRC and K.F.L. by Amgen Inc.

Sequence of *C. elegans lag-2* reveals a cell-signalling domain shared with *Delta* and *Serrate* of *Drosophila*

Frans E. Tax, Jon J. Yeagers & James H. Thomas

Department of Genetics, University of Washington, Seattle, Washington 98195, USA

THE *lin-12* and *glp-1* genes of *Caenorhabditis elegans* encode members of the *Notch* family of transmembrane proteins^{1,2}. Genetic studies indicate that the *lin-12* and *glp-1* proteins act as receptors in specific developmental cell interactions^{3–6} and that their functions are partially redundant⁷. *lin-12 glp-1* double mutants display certain embryonic defects not found in either single mutant^{7,8}. The phenotype of this double mutant is called *Lag*, and recessive mutations in either of the genes *lag-1* or *lag-2* can also result in the *Lag* phenotype⁷, indicating that these two genes may participate in the same cell interactions that require *lin-12* or *glp-1*. We report here that *lag-2* encodes a predicted transmembrane protein of 402 amino acids. The predicted extracellular region of *lag-2* is similar to amino-terminal regions of *Delta* and *Serrate*, two *Drosophila* proteins that are thought to function as ligands for *Notch*^{9–14}. The region of similarity includes sequences related to epidermal growth factor (EGF) repeats. We have isolated *lag-2(sa37)*, a dominant allele that shows specific genetic interactions with *lin-12*. The *sa37* mutation causes a Gly→Asp change in a conserved residue of an EGF motif. Because of its overall structure, its sequence similarity to *Delta* and *Serrate*, and its genetic interactions, we suggest that *lag-2* encodes an intercellular signal for the *lin-12* and *glp-1* receptors.

We cloned *lag-2* by mapping chromosomal deficiencies to align the physical and genetic maps in the *lag-2* region (Fig. 1a, b) and by using germ-line transformation to identify the relevant DNA (Fig. 1b, c). A 3.8-kilobase fragment of genomic DNA (pTJ225) was sufficient to rescue all *lag-2* loss-of-function phenotypes. A portion of this fragment was used as a probe

TABLE 1 Quantification of suppression by *lag-2(sa37)* in *lin-12(d)* strains

<i>lag-2</i> genotype	<i>lin-12(dom)</i> genotype	Per cent non-Vul	Number
+/+	+/+	100	>500
+/+	<i>n302/n302</i>	0	>500
+/+	<i>n952/n952</i>	0	>500
+/+	<i>n950/n950</i>	0	>500
<i>sa37</i> /+	<i>n302/n302</i>	27	109
<i>sa37/sa37</i>	<i>n302/n302</i>	96	351
<i>sa37/sa37</i>	<i>n952/n952</i>	90	214
<i>sa37/sa37</i>	<i>n950/n950</i>	35	255
<i>sa37/sa37</i>	+/+	100	>500

lin-12(d) alleles are in order of increasing strength, as determined by the penetrance of other mutant defects³. *lag-2(sa37)* was isolated after ethylmethanesulphonate (EMS) mutagenesis as a suppressor of the vulvaless (Vul) phenotype of *lin-12(n950)* (ref. 3). *sa37* was mapped to the left arm of linkage group V (LGVL) to less than 0.2 map units from another dominant *lin-12* suppressor, *lag-2(n1255)* (J.H.T., F.E.T. and H.R. Horvitz, manuscript in preparation). To quantify suppression, worms were picked to individual plates and allowed to lay eggs for 1–2 days, or until they formed a bag of worms (for Vul animals). All progeny from each parent were inspected with a dissecting microscope for the presence of a normal vulva. Presence of a vulva indicates suppression. *sa37*/+, *sa37/deficiency* and *sa37/sa37* animals in both *lin-12(+)* and (d) strains were examined using a dissecting microscope and no additional defects were observed.

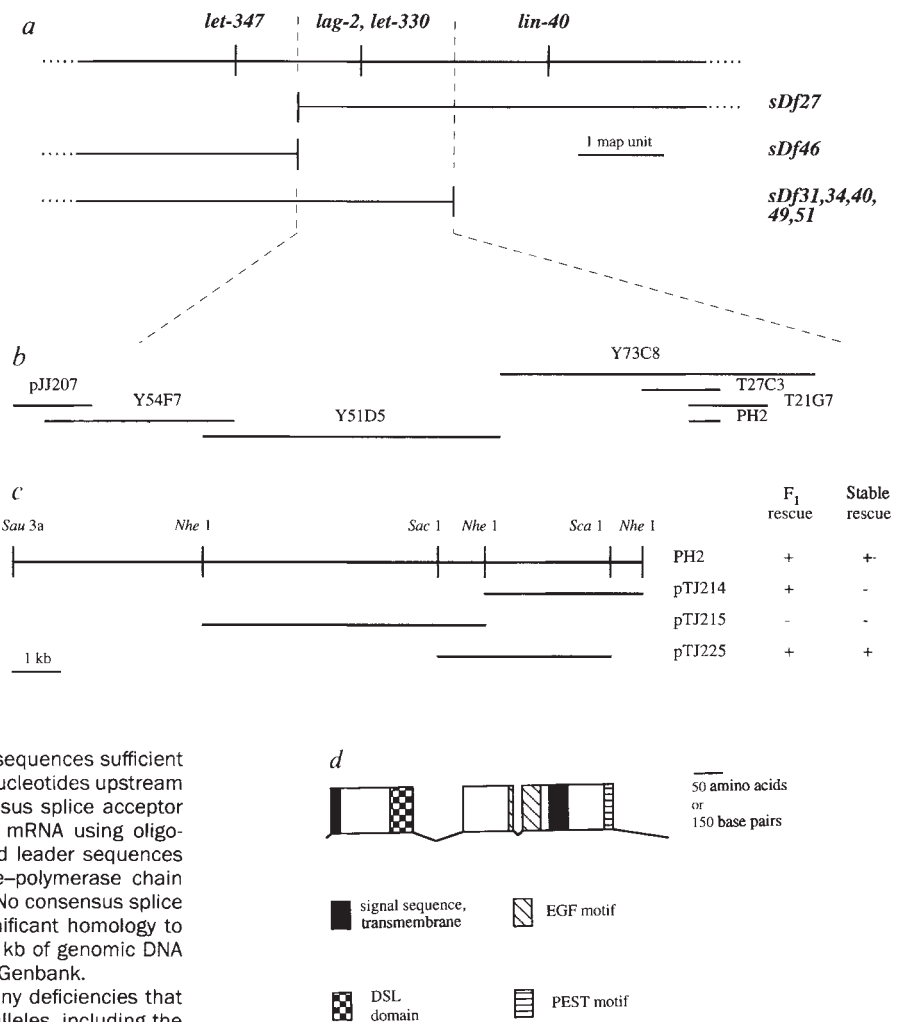
to isolate complementary DNAs. One cDNA and most of the genomic clone were sequenced (Fig. 2). The *lag-2* cDNA is predicted to encode a protein of 402 amino acids with a potential signal sequence and one potential transmembrane domain (Figs 1d and 2). Protein database searches showed significant similarity between *lag-2* and the extracellular domains of two *Drosophila* genes: a segment of the extracellular domain of *lag-2* shares 23% identity with a 187-amino-acid region of *Delta*^{9,10} and 23% identity with a 229-amino-acid region of *Serrate*^{11,12} (Fig. 3a). This region is a subset of the amino-terminal segment found to be most similar between *Delta* and *Serrate*¹². Of the 45 amino-acid identities in the region of *lag-2* and *Delta* similarity, 30 are also conserved in *Serrate*. The extracellular domain of *lag-2* contains several clusters of cysteines and other residues that are conserved in the EGF motif¹⁵ (Fig. 3b, c). This motif is found

in many signalling molecules and receptors including EGF, *Notch*, *lin-12* and *glp-1* (refs 1, 15, 16), often in many tandem repeats. Adjacent to the putative transmembrane domain is a fully conserved EGF repeat with cysteine spacing identical to certain EGF repeats from *glp-1* (ref. 1) and *Notch*¹⁶ (Fig. 3c, d). This EGF repeat is also very similar to the third full repeat of *Delta* and *Serrate* (Fig. 3c), differing in the spacing of cysteines only by one gap of a single amino acid. The predicted intracellular domain of *lag-2* contains a potential PEST sequence (single-letter amino-acid code), which is common among proteins with short half-lives¹⁷. No other regions of *lag-2* showed significant similarity to other proteins.

The region of extended similarity between *lag-2*, *Delta* and *Serrate* contains what has been called a 'partial' or 'degenerate' EGF repeat because it only weakly matches the consensus¹¹.

FIG. 1 Genetic and molecular maps of *lag-2*.

a, A genetic map showing the position of *lag-2* on LGVL. The extent of deficiencies is indicated by solid lines²⁴. *let-330* and *lag-2* have been mapped to the same deficiency interval but have not been ordered²⁴. **b**, Physical map of the region between the left breakpoint of *sDf27* and the right breakpoints of *sDf31*, *sDf34*, *sDf40*, *sDf49* and *sDf51*, as determined by dosage studies of deficiency-containing strains (see Methods). Clones beginning with a Y are YAC clones. pJJ207 is a plasmid subclone of *srg-1* (J. Priess, personal communication). PH2 is a phage clone (see Methods). All others are cosmid clones, and only relevant cosmids are shown. **c**, Molecular map of the immediate *lag-2* region. The ability of genomic clones to rescue the recessive lethal *lag-2* allele *n1255 n1323* is tabulated on the right. **d**, An expanded view of the *lag-2* transcription unit. The two introns and 5' and 3' untranslated sequences are shown as lines; exons are boxed. The cDNA contains a poly(A) tail which is not present in the genomic DNA and is flanked on each side by at least 0.5 kb of genomic DNA in pTJ225, the fully rescuing plasmid. pTJ214, which contains 115 bp of genomic DNA upstream of the first in-frame methionine, gave F₁ transgenic rescue, but all such animals were sterile. This suggests that all necessary *lag-2* coding sequences were present, but not promoter sequences sufficient for germ-line rescue of *lag-2* phenotypes. At three nucleotides upstream of the first in-frame methionine, there is a consensus splice acceptor site. Attempts to identify the 5' end of the *lag-2* mRNA using oligo-nucleotide primers for the *C. elegans* trans-spliced leader sequences (SL1 and SL2) in controlled reverse transcriptase-polymerase chain reaction (RT-PCR) experiments were unsuccessful. No consensus splice donor sites and no open reading frames with significant homology to protein databases were found in the upstream 1.5 kb of genomic DNA in pTJ225. All sequences have been submitted to Genbank. **METHODS**. Complementation analysis: *let-330*, many deficiencies that map to LGVL²⁴, and several *lag-2* loss-of-function alleles, including the first-reported allele called *let-461(s1486)* (ref. 24), were analysed (refs 7, 24, and J.H.T., F.E.T., and H. R. Horvitz, manuscript in preparation). Complementation tests were done by assessing the viability of lethal and deficiency alleles in *trans* heterozygotes. Additional alleles of *lag-2* have been found⁷ which also failed to complement *lag-2(s1486)*. Alignment of physical and genetic maps: Deficiencies for LGVL were balanced using a lethal-carrying form of the translocation *eT1* (ref. 25, and M. Wakarchuk and D. Baillie, personal communication). Only deficiency heterozygotes were viable in these strains. A set of dilutions of restriction-enzyme-digested genomic DNA was run with DNA from deficiency heterozygotes. Southern blots were prepared as described²⁶. Whole cosmids or DNA fragments from selected cosmids were used as probe. Band intensities on autoradiograms were assessed to infer the map position of *lag-2* shown in **a** and **b** (data not shown). Transformation experiments: Purified DNA from candidate cosmids and pRF4 (ref. 27; a transformation marker) was microinjected into *lag-2(n1255 n1323)*



unc-46(e177); *yDp1* animals. *yDp1* (*IV;Vf*) (ref. 28) is a duplication that carries both *lag-2*(+) and *unc-46*(+). Rescue of *lag-2* was scored only by the presence of viable *unc-46* animals both in the F₁ (transient rescue) and in the F₃ from animals expressing the transformation marker in the F₂ (stable rescue). Both cosmids T21G7 and T27C3 stably rescued *lag-2*. A set of genomic λ clones (T. Blumenthal, personal communication) was isolated using the cosmid overlap as probe. Fragments from a phage (PH2), which stably rescued *lag-2*, were cloned into Bluescript plasmids (Stratagene) and tested for rescue (1c). Approximately 30% of the progeny from a self-fertilizing hermaphrodite carrying a heritable line were viable and fertile. We did not observe any new phenotypes in these rescued lines. Restriction digests, cloning, plasmid and cosmid DNA isolation, and plaque screening followed standard methods²⁶.

However, the amino-acid similarity and conserved spacing of the cysteines in this region suggest that this sequence represents a specific variant of the EGF repeat with functional significance. Because of its conserved sequence, we name the region shown in Fig. 3b the DSL domain, for *Delta*, *Serrate* and *lag-2*. The functional significance of the DSL domain is supported by *in vitro* cell-aggregation assay for *Notch* binding^{13,14}; analyses of *Delta* and *Serrate* demonstrate that this region is part of the sequence required to bind *Notch* protein (S. Shepard and M. Muskavitch, and R. Fleming, personal communications).

We have isolated a new dominant allele of *lag-2*, *sa37*, which suppresses *lin-12(d)* alleles (Table 1). We have shown that *sa37* is a *lag-2* allele on the basis of its genetic map position, its specific interaction with *lin-12*, and its DNA sequence (Table 1; Fig.

3c). The *lin-12(d)* alleles, proposed to encode interaction-independent receptors¹⁸, have been ordered in an allelic series on the basis of the penetrance of various mutant defects³. *sa37* suppresses all tested *lin-12(d)* alleles, showing that suppression is not allele-specific (Table 1). As the *lag-2(sa37)* single mutant has no discernible mutant phenotype (Table 1), the allele must retain some wild-type *lag-2* function. We sequenced the coding region of the dominant *sa37* allele of *lag-2* and found one change in a conserved residue of the EGF motif (glycine at residue 262 to aspartate) (Fig. 3c, d). The *lin-12(d)* alleles are missense mutations in the extracellular region of *lin-12* (ref. 18). We hypothesize that *lag-2(sa37)* causes an altered signal that affects the function of the *lin-12(d)* receptor. Interestingly, certain dominant *Notch* alleles are *Delta*-dependent¹⁹.

FIG. 2 DNA sequence of *lag-2*. Nucleotide and predicted amino-acid sequences corresponding to the *lag-2* cDNA and genomic DNA. Amino acids are numbered in bold face starting with the first predicted methionine. The amber termination codon is denoted by an asterisk. Consensus splice acceptor and donor sites are overlined. The putative PEST sequence is bracketed. The nucleotides indicating the start of the cDNA and the site for the addition of the poly-A tail in the genomic DNA are indicated with dots above them. The putative signal sequence and TM are underlined with a dashed line. There are two introns. A failure to splice either intron would terminate the protein at the first splice junction or three amino acids past the second splice junction, resulting in small secreted proteins lacking a TM. Alternative splicing could result in a smaller, in-frame protein lacking 73 amino acids of the extracellular domain. There are also several potential protease cleavage sites²⁹ in the extracellular domain, indicated with carets, which could be used to liberate secreted peptides. We have no evidence that any of these alternative forms of the protein are made *in vivo*.

METHODS. cDNA cloning and sequencing: The genomic DNA insert of pTJ214 was used as probe to screen two *C. elegans* cDNA libraries (P. Okkema, V. Plunger and A. Fire, personal communication). Two identical cDNA clones were found among ~300,000 plaques from an embryonic library and no clones were found among ~200,000 plaques from a mixed-stage library. One cDNA insert was cloned into a Bluescript vector to generate deletions for sequencing. Double-stranded plasmid DNAs were sequenced using Sequenase (US Biochemical) by standard methods³⁰. Gaps left in this sequence were crossed using specific oligonucleotides as primers. Both strands of the cDNA were sequenced, and portions of several genomic plasmids were sequenced.

```

ctagccaaaactttcactgtgtcttttttctcctaagtattttccgaacttatagtc 60
gaccttgagcttgctagaagccaacaacacacttttcaaatgtgctttttcagaaa 120
atgatcgcttacttctcttactctcactgcctcccgactctccaagccgagtagaa 180
M I A Y F L L L L T C L P V L Q A R V E 20
gtccatcaagagttcactctctcgaagcaggtctccgtccgttcgaattgtcacagag 240
V H Q E F I S S K R V S V R F E I V T E 40
tccactcgccgaatgtccagtcacatttgacctcttccccgcggcccaaaccaac 300
S H S P N R P V T F D L P R G P K T N 60
atcatctctctcgacacgttcaaccgggtcttcaacttttccattcaactgtgcagccg 360
I I L L D T F N P V F N F S I Q L V Q P 80
ttcactggcgagccgctcggtgatcggtatctatcgaaagttcagttttcgggcacaa 420
F T G Q P L G D R I Y R K V Q F S G T N 100
cagcgtgatcaacgacggttcacgacgacatctggaattctgtgagcgtggcgacc 480
Q P W I N D T F T T T S G I S L S V A T 120
gaggtcacatgtgctcgcaactatttcgggaatcgatcgcaaaacttttggatgcacac 540
E V T C A R N Y F G N R C E N F C D A H 140
ttggcgaaggcgcgaggaagagatgcgacgcatgggagcgtctccgtgtgacatcgga 600
L A K A A R K R C D A M G R L R C D I G 160
tggtgggacacattgtggcagaggtgagctttaacttcagtggtgtttacgtggtat 660
W M G P H C G Q 168
cagcgtgtcccatcaggttgatctacaaaaaaatgcggaattttttggctcaaaaatg 720
tgacgtcagcaggttcttaactgggtctcaactcccgcatttttttgtagatctaggt 780
gatcaaacggaaatgggacactctggcaccacgtgtgtgtctcaaaactctctatagatt 840
ttttcaagtggttttaatttcagcgttgaccccgcaagtgctcatgagcagaacgagaa 900
A V D P R K C S C E N D G 181
tctgcgtctcgtgatgattcatcaagtcagcgaatcaaacatcgagcaatgaacagt 960
I C V S C M I H P S Q P N Q T S S N E C 201
tgatctgcgaatgcaccaacgggttcacggcgttgctggcaatttttcggtctcaatc 1020
L I C E C T N G P T G T R C E I F G F N 221
agttccagctgacccgccccgacagatgcttgcagtgctcaagacgctgcttgaatg 1080
Q F Q L T A P R P D A C S V K D A C L N 241
gttagtgtttttgttagagtcagacatgagaattgttttttcagcgcgaatgtcttc 1140
G A K C F 246
ccaaacggctcgaagttcttctgctcgtcgtgctggttccgtggaattctgtgaa 1200
P N G P K V F C S C A V G F I G E F C E 266
atctcactcaccaccaccaccacacacacacacacacacacacacacacacacacac 1260
I S L T T T T P T T V E I T V T S S G Y 286
tccagcgcgtctacatcacagttgctctctctgctgattctctcatcatcctggatgc 1320
S S A V Y I T V A L F V I F S I I I G C 306
ttcaagtacaagttcaagcagatgagacagcaggtctggcctgtggacaggtaccagag 1380
F K Y K F K P M R Q Q A L A R G Q V P E 326
ccctacaagatgctgagacaaagtcagtgatgctgacgacgggaagctcagaggtcag 1440
P Y K M P E T K S M L I D P E A S E A Q 346
aagaaggtgttcacatcgagggcagtgatgcagaagattgaggaagtgaggtatacc 1500
K K V F T I E G S V Q K I D E E V R Y T 366
agtgtcccaaggaagtgaatcgaataatgagtgatgccgtttatcaaaagctactcca 1560
S A P R K Y E S N N E Y A V I Q [K S T P 386
ccgcaccattcttaagtcacgtctcttccagcctgtcactgtcattgtcagattcttct 1620
P P P S L S P P S I P C A H Y V * 402
ttatccaaaaatgtgtgtctcttccacagtaaaaaattttccaggttttttctcagtt 1680
tctctctcttatctcattttctttaaactgtttccctctgtgattgctgtaatttcga 1740
tctcctacattccgtccatctccacacctctccagcatcccggttctgttttctattt 1800
cattgtattctgaaatgtttgtgtaatttatactttttatactttttgtttctgtttt 1860
tgaaaaaaatgtgatccaaatcttgaaatctggtccacacgtgtgtgacaggtgca 1920
gtgtgaaagcgcaacaaatgatacgaaaaattttccatcaaaaaattggcaaaaag 1980

```

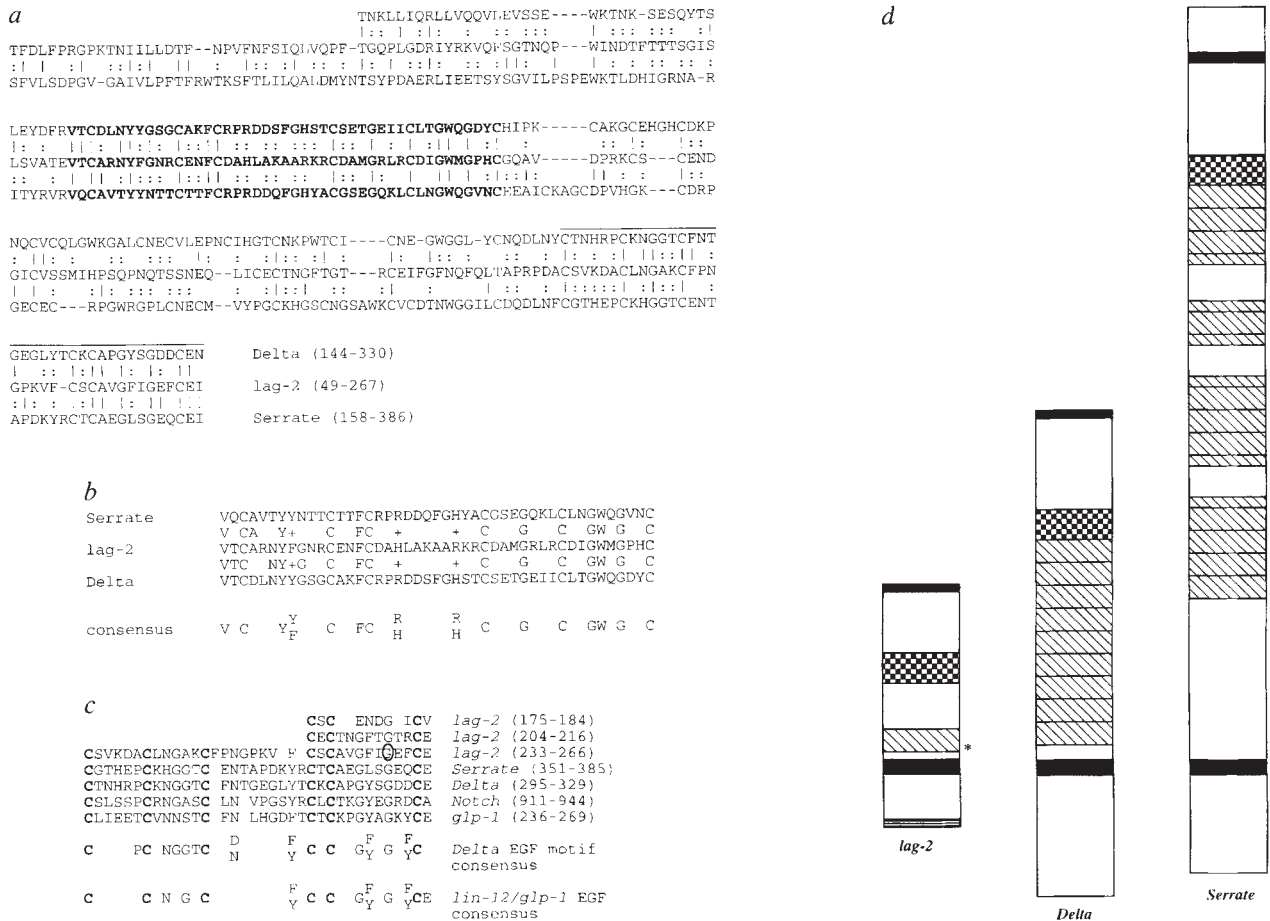


FIG. 3 Sequence comparisons of *lag-2*. *a*, Alignment showing most of the extracellular region of *lag-2* with N-terminal regions of *Delta*¹⁰ and *Serrate*¹¹. The alignment is arranged to show maximum similarity of *Delta* and *Serrate* to *lag-2* and does not give the best alignment for *Delta* and *Serrate* in this region¹². Total identities are *lag-2-Delta* 45/191, *lag-2-Serrate* 55/234 and *Delta-Serrate* 88/191. The DSL domain (b) is in bold type and the EGF motif (c) is overlined. For both the DSL region and the longer region of homology, alignment scores for randomized *lag-2* sequences were much lower than the scores of *lag-2* with *Delta* and *Serrate* (differences were significant at $P < 0.001$ by the Mann-Whitney *U* test). *b*, DSL domain of *lag-2*, *Delta* and *Serrate*. A consensus was derived by noting either identical or very conservative residues in all three proteins. We have searched Genbank, SWISS-PROT and PIR (ref. 31) for this conserved spacing of cysteines and no other close matches to the consensus were found. Total identities for the DSL domain are *lag-2-Delta* 16/45, *lag-2-Serrate* 14/45, *Delta-Serrate* 24/45. Of the 24 identities between *Delta* and *Serrate*, 13 are conserved in *lag-2*. *c*, An alignment of three regions of *lag-2* with other EGF motifs and a consensus derived from *Delta*¹⁰ and from *lin-12* and *glp-1* (ref. 1). Cysteines are indicated in bold type. The sequences shown are EGF 23 from *Notch*¹⁶, EGF 5 from *glp-1* (ref. 1) and EGF 3 from *Delta*^{9,10} and *Serrate*^{11,12}. The circled glycine is altered by the mutation

On the basis of these observations, we suggest that *lag-2* is a ligand for both *lin-12* and *glp-1*. The signals for *lin-12* and *glp-1* are similar or identical²⁰, and if there is only one signal for these receptors, recessive loss-of-function mutations in that gene should cause the Lag phenotype. Although we cannot rule out other possibilities, such as a function as a transmembrane co-receptor with *lin-12* and *glp-1*, the sequence of *lag-2* is most consistent with action as a ligand: the predicted protein contains sequences similar to the EGF motif, and is similar to *Delta* and *Serrate*, which are probably *Notch* ligands in *Drosophila*.

Homologues of *Notch* are found in both vertebrates and invertebrates^{1,21}. Genetic analysis of *C. elegans* and *Drosophila melanogaster* shows that *Notch* and its homologues function as

in *lag-2(sa37)*: Gly to Asp, GGT to GAT. *d*, A visual comparison of the domains of *lag-2*, *Delta* and *Serrate*. The pattern for motifs is the same as in Fig. 1*d*. The position of the lesion in *sa37* is indicated by the asterisk.

METHODS. Sequence of *lag-2(sa37)*: Two oligonucleotide primers (5'GGATAAAGGAAGAAATCTAGAC3' and 5'CGCTTACTTCCTCTTACTCC3') were used to amplify the *lag-2* coding region, and the sequencing primers in Fig. 2 and additional primers were used to sequence the amplified DNA. Templates were sequenced using the Applied Biosystems Model 373A system. Both strands of *lag-2(sa37)* were sequenced and the indicated change was the only one seen. Database search analysis: The alignment shown in a was found by searching the EMBL protein database using TFASTA. The *lag-2* amino-acid sequence was analysed using the Motifs program from GCG, BLAST from NCBI, and FASTDB from IntelliGenetics. The motifs program detected two EGF motifs in the *lag-2* coding sequence, at amino acids 204–215 and 233–265. The three N-terminal cysteines that define the EGF motif are not present before amino acids 204–215 (c), so we do not consider this to be an EGF motif. Amino acids 175–183 did not score as an EGF motif, but the pattern of three cysteines resembles the C-terminal pattern found in the EGF motif and the pattern seen in amino acids 204–215.

receptors during cell interactions^{3 6,22}. Mosaic analysis of *Delta* in *Drosophila* indicates that *Delta* acts as a signal²² and analysis of *Delta* protein expression²³ is consistent with this view. There have been no previous reports of genes similar to *Delta* or *Serrate* in other species. The sequence of *lag-2* suggests that ligands of *Notch*-family receptors will also be widely conserved and will share the DSL domain and at least one consensus EGF motif. □

Received 2 November; accepted 13 December 1993.

1. Yochem, J. & Greenwald, I. *Cell* **58**, 553-563 (1989).
2. Austin, J. & Kimble, J. *Cell* **58**, 565-571 (1989).

3. Greenwald, I. S., Sternberg, P. W. & Horvitz, H. R. *Cell* **34**, 435–444 (1983).
4. Priess, J. R., Schnabel, H. & Schnabel, R. *Cell* **51**, 601–611 (1987).
5. Austin, J. & Kimble, J. *Cell* **51**, 589–599 (1987).
6. Seydoux, G. & Greenwald, I. *Cell* **57**, 1237–1245 (1989).
7. Lambie, E. J. & Kimble, J. *Development* **112**, 231–240 (1991).
8. Bowerman, B., Tax, F. E., Thomas, J. H. & Priess, J. R. *Development* **116**, 1113–1122 (1992).
9. Vassini, H., Bremer, K. A., Knust, E. & Campos-Ortega, J. A. *EMBO J.* **6**, 3431–3440 (1987).
10. Kocyński, C. C., Alton, A. K., Fechtel, K., Kooh, P. J. & Muskavitch, M. A. T. *Genes Dev.* **2**, 1723–1735 (1988).
11. Fleming, R. J., Scottgale, T. N., Diederich, R. J. & Artavanis-Tsakonas, S. *Genes Dev.* **4**, 2188–2201 (1990).
12. Thomas, U., Speicher, S. A. & Knust, E. *Development* **111**, 749–761 (1991).
13. Fehon, R. G. et al. *Cell* **61**, 523–534 (1990).
14. Rebay, I. et al. *Cell* **67**, 687–699 (1991).
15. Appella, E., Weber, I. T. & Blasi, F. *FEBS Lett.* **231**, 1–4 (1988).
16. Wharton, K. A., Johansen, K. M., Xu, T. & Artavanis-Tsakonas, S. *Cell* **43**, 567–581 (1985).
17. Rogers, S., Wells, R. & Reichsteiner, M. *Science* **234**, 364–368 (1986).
18. Greenwald, I. & Seydoux, G. *Nature* **346**, 197–199 (1990).
19. Heitzler, P. & Simpson, P. *Development* **117**, 1113–1123 (1993).
20. Seydoux, G., Schedl, T. & Greenwald, I. *Cell* **51**, 939–951 (1990).
21. Coffman, C., Harris, W. & Kintner, C. *Science* **249**, 1438–1441 (1990).
22. Heitzler, P. & Simpson, P. *Cell* **64**, 1083–1092 (1991).
23. Kooh, P. J., Fehon, R. G. & Muskavitch, M. A. T. *Development* **117**, 493–507 (1993).
24. Johnson, R. C. & Baillie, D. L. *Genetics* **129**, 735–752 (1991).
25. Rosenbluth, R. E. & Baillie, D. L. *Genetics* **99**, 415–428 (1981).
26. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, New York, 1989).
27. Mello, C. C., Kramer, J. M., Stinchcomb, D. T. & Ambros, V. *EMBO J.* **10**, 3959–3970 (1991).
28. DeLong, L., Casson, L. D. & Meyer, B. J. *Genetics* **117**, 657–670 (1987).
29. Loh, Y. P., Brownstein, M. J. & Gainer, H. A. *Rev. Neurosci.* **7**, 189–222 (1984).
30. Sanger, F., Nicklen, S. & Coulson, A. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
31. Altschul, S. F. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5509–5513 (1990).

ACKNOWLEDGEMENTS. We thank H. R. Horvitz for support of initial analysis of *lag-2* genetics; A. Coulson and J. Sulston for cosmids; C. Mello for help with transformation; R. Johnson and D. Baillie for lethals and deficiency strains; P. Okkema, V. Plunger and A. Fire for cDNA libraries; B. Bowerman for purified mRNA; T. Blumenthal for the genomic library; L. Nagy and S. Henikoff for help with homology searches; L. Nagy, D. Reiner, C. Mello, C. Goutte, B. Draper, J. Priess and B. Bowerman for comments on the manuscript; and especially J. Priess and members of our laboratories for discussion and assistance.

The human X-inactivation centre is not required for maintenance of X-chromosome inactivation

Carolyn J. Brown & Huntington F. Willard

Department of Genetics, Center for Human Genetics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA

X-CHROMOSOME inactivation occurs early in mammalian female development to achieve dosage compensation with males¹. Although it is widely accepted that this inactivation requires the presence in *cis* of the X-inactivation centre (XIC)^{2–5}, it is not known whether the XIC is required for the initiation, promulgation or maintenance of X inactivation^{6,7}. The *XIST* gene, which is localized within the XIC interval on both the human and mouse X chromosomes, is constitutively expressed from inactive X chromosomes^{8–10}, suggesting a possible role in the maintenance of X inactivation. To address whether the presence of the XIC, including the *XIST* gene, is continuously required for the maintenance of X-chromosome inactivation, we have analysed the transcriptional activity of a number of X-linked genes in mouse/human somatic cell hybrids retaining an intact human inactive X chromosome or derivatives of the inactive X chromosome lacking the XIC. Genes subject to X inactivation remain transcriptionally silent despite the loss of the XIC, demonstrating that the presence of the XIC is not required for the maintenance of X inactivation in somatic cells.

A series of mouse/human somatic cell hybrids which retain the human inactive X chromosome with or without the XIC region was isolated. More than 30 hybrids (called 'pushmi-pullyu' hybrids after the counter-selection technique summarized in Fig. 1) were isolated and analysed for markers between *UBE1* (in Xp11.23) and the XIC (in Xq13.2) by Southern blot and polymerase chain reaction (PCR) analysis. Table 1 outlines the results from PCR analysis of a number of the hybrids, showing that the breakpoints are distributed on the long arm of the X chromosome proximal to the XIC. More detailed analysis of the breakpoints in several of these hybrids has shown that the breakpoints are independently derived and that no complex rearrangements have occurred¹¹. As the XIC has been mapped between *DXS227* and *DXS56* (ref. 5), and as both these loci, as well as the *XIST* gene, are missing in all of the inactive X pushmi-pullyu clones, we conclude that all these hybrids have lost the XIC interval.

To determine whether the loss of the XIC has any effect on maintenance of X-chromosome inactivation, the transcription of several genes retained in the pushmi-pullyu hybrids was analysed by reverse transcriptase (RT)-PCR (Fig. 2). Primers for the *MIC2* cell-surface antigen gene (expressed from both the active and inactive X chromosomes¹²) amplified product in all hybrids. However, no expression of the *POLA* (DNA polymerase- α) or *PDHAI* (α -subunit of pyruvate dehydrogenase) genes was detected in any of the pushmi-pullyu hybrids, demonstrating that the loss of the XIC did not result in reactivation of either of these genes. As expected, no expression of the *XIST* gene was observed from the pushmi-pullyu hybrids. In addition, the mouse *XIST* gene was not expressed in either the parental mouse cell line or the pushmi-pullyu hybrids, eliminating the possibility of an interspecies effect on X-chromosome inactivation.

TABLE 1 Retention of proximal Xq markers in the pushmi-pullyu hybrids

Cell line	t11...	t11pp	t11pp	t11pp	t11pp	t11pp	t11...	tpp	t11pp	t11pp	t11pp	t11pp	t86...	t86pp	t86pp
Locus	10d*	4	B1	C1	D1	D4	2b*	1a	4a	5a	6a	11a	B2*	6	18
DXZ1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
DXS136	+	–	+	+	+	–	+	–	+	+	+	+	+	–	–
DXS339	+	–	–	–	+	–	+	–	–	+	+	+	+	–	–
DXS227	+	–	–	–	–	–	+	–	–	–	–	–	+	–	–
XIST	+	–	–	–	–	–	+	–	–	–	–	–	+	–	–
DXS56	+	–	–	–	–	–	+	–	–	–	–	–	+	–	–

The XIC interval is distal to *DXS227* and proximal to *DXS56* and contains the *XIST* gene⁵. As previously seen for active X-chromosome pushmi-pullyu hybrids¹⁸, all hybrids retained α -satellite (*DXZ1*) at the human X-chromosome centromere. Plus and minus signs designate amplification results from PCR using the following primers under standard conditions²³: *DXZ1*: 3A, ATAATTCCATACTAAACACA; 4A, TGTGAAGATAAAG-GAAAAGGCTT. *DXS136*: 1, CAGCTCAATGCTGGAGAAGA; 2, GGGAAGGTGTACCCTTCA. *DXS339*: 1, TGGCCATTTTCATCACCAGT; 7, GGTCATTGGGTCCAGGCTT. *DXS227*: A, TATCAGAACATGGTCAGTT; B, CATGGGTGCCTGGTTTGTGTT. *XIST*: 1, GAAGTCTCAAGGCTTGAGTTAGTAGAAG; 2r, TTGGGTCCTCTATCCATCTAGGTAG. *DXS56*: 1, GTGGGATTACTGTATTATTCAGA; 2, ATACTCAGTAGCCACTACTC.

* The parental hybrid lines with an intact inactive X chromosome before pushmi-pullyu selection. The next columns show their pushmi-pullyu derivatives (names with pp). Full cell-line designations are: t11...10d=t11-4Aaz5-Hat1-10d; t11...2b=t11-4Aaz5ac2b; and t86...B2=t86-B1maz1b-3a2ac1bB2.

TABLE 2 Summary of expression of X-linked genes in somatic cell hybrids with and without the XIC

Gene	<i>MIC2</i>	<i>PRPS2</i>	<i>PDHA1</i>	<i>POLA</i>	<i>ARAF1</i>	<i>TIMP1</i>	<i>ELK1</i>	<i>TFE3</i>	<i>AR</i>	<i>XIST</i>
Cell line										
t11...10d	+	—	—	—	—	—	—	—	—	+
t11pp4	+	—	—	—	—	—	—	—	DEL	DEL
t11ppB1	+	—	—	—	—	—	—	—	—	DEL
t11ppC1	+	—	—	—	—	—	—	—	—	DEL
t11ppD1	+	—	—	—	—	—	—	—	—	DEL
t11...2b	+	—	—	—	—	—	—	—	—	+
t11pp1a	+	—	—	—	—	—	—	—	DEL	DEL
t11pp5a	+	—	—	—	—	—	—	—	—	DEL
t11pp6a	+	—	—	—	—	—	—	—	—	DEL

For columns marked DEL, the hybrid is deleted for that gene. All the genes are subject to X inactivation except *MIC2*, which is expressed from both active and inactive X chromosomes¹², and *XIST*, which is expressed only from inactive X chromosomes⁸. The cell line designations are as in Table 1, and RT-PCR was performed with human-specific primers as listed in Fig. 2 legend as well as for *TIMP23* and the following primer pairs derived from published sequences: *PRPS2*: 5'-CTCCGCCACCTCCTCCGC-3'; 5'-TGATGAGGAGTTCATCAGG-3'. *ARAF1*: 5'-TGGAGATGGAGGAGCTCCCA-3'; 5'-TCAGCAAAATCTCCAGCAAC-3'. *ELK1*: 5'-GGACCTAGAGCTTCCACTCA-3'; 5'-AGAGCATGGATGGAGTGACC-3'. *TFE3*: 5'-GCTGGAGAAGGACATGAGAT-3'; 5'-GACGGACTAGACTGGTGGTT-3'. *AR*: 5'-AGGAAAGCGACTTCACCGCA-3'; 5'-GAGCTCCATAGTGACACCA-3'.

Expression of additional X-linked genes known to be subject to X inactivation was also examined, as genes may be differentially influenced by factors involved in the maintenance of X-chromosome inactivation dependent, for example, on the presence or strength of a CpG island subject to DNA methylation or proximity to replication origins^{6,13-15}. No reactivation was observed for any of the genes analysed (Table 2), including seven from the short arm of the X chromosome and one from its

proximal long arm. Expression of all these genes has been detected from somatic cell hybrids retaining the active X chromosome. That the continued inactivation of the genes examined was due to maintenance of X-chromosome inactivation, rather than a position effect caused by association with mouse chromosomes in the hybrids, was demonstrated by observing the continued expression of a gene in the proximal X short arm which escapes X-chromosome inactivation, and by fluorescence

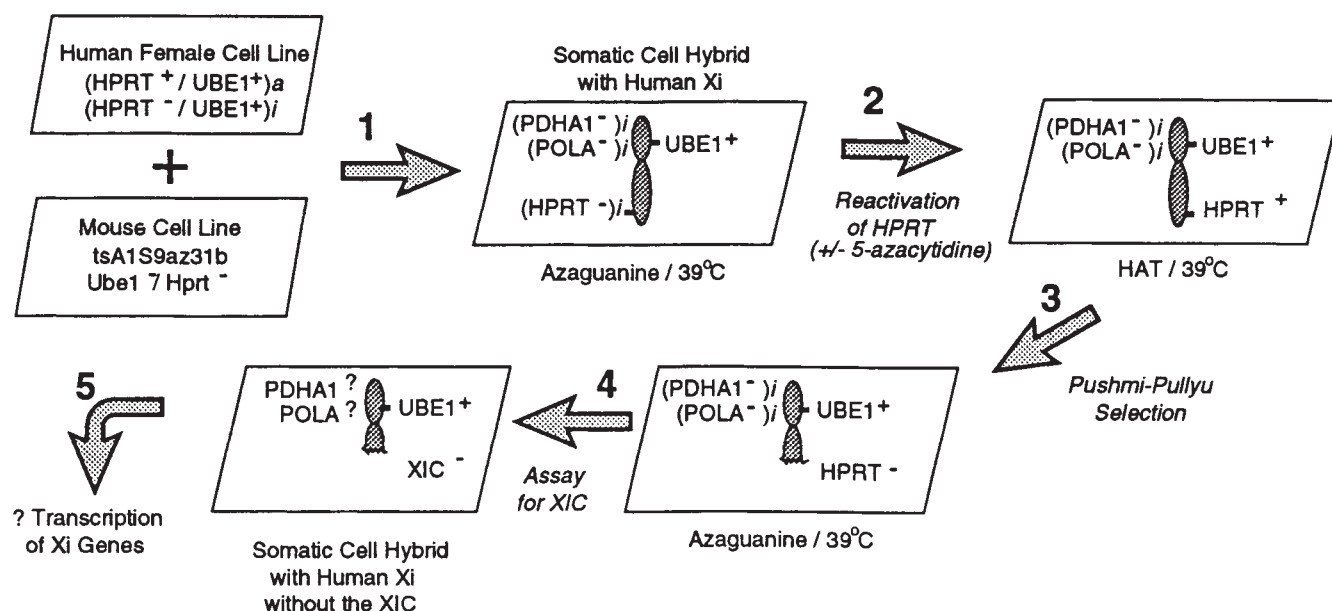


FIG. 1 Diagram of selection for pushmi-pullyu hybrids¹⁸ retaining the inactive X chromosome without the XIC. Stage 1: Human female cell lines containing an inactive X chromosome were fused with the mouse tsA1S9az31b cell line which has a temperature-sensitive defect in the *ube1* gene that can be complemented by either the active or inactive human X chromosome²⁴. Hybrids which had lost the human active X, but retained the inactive X, were obtained after growth at the non-permissive temperature (39 °C) in medium containing azaguanine and thioguanine (AGTG). The presence of the inactive X chromosome was confirmed enzymatically, cytogenetically and by RT-PCR or northern analysis. Stage 2: The presence of two selectable markers on a single chromosome makes it possible to select for breakpoints between the two markers²⁵, and we have previously used selection for the human *UBE1* gene in combination with selection against the *HPRT* gene to generate a series of hybrids with X chromosomal breakpoints between

the two markers on active X chromosomes¹⁸. To use a similar selection scheme with the inactive X chromosome, it was necessary to isolate reactivants of the *HPRT* gene on the inactive X chromosome, either spontaneously (t11-4az5-Hat1-10d) or after treatment with the demethylating agent 5-azacytidine (t86-B1maz1b-3a2ac1bB2 and t11-4Aaz5ac2b) by selection in HAT (hypoxanthine-aminopterin-thymidine) medium. Stage 3: The pushmi-pullyu selection conditions (AGTG and 39°C) selected for those chromosomes which had broken between the *UBE1* and *HPRT* genes, losing *HPRT* enzyme activity. *HPRT*⁻/*UBE1*⁺ colonies were isolated at a frequency of ~10⁻⁴. Stage 4: Analysis by PCR for the presence of the XIC showed that all clones isolated had lost the XIC. (5) XIC⁻ clones were then analysed for the expression of genes subject to X inactivation by RT-PCR analysis, to test the effect of the loss of the XIC on maintenance of X-chromosome inactivation.

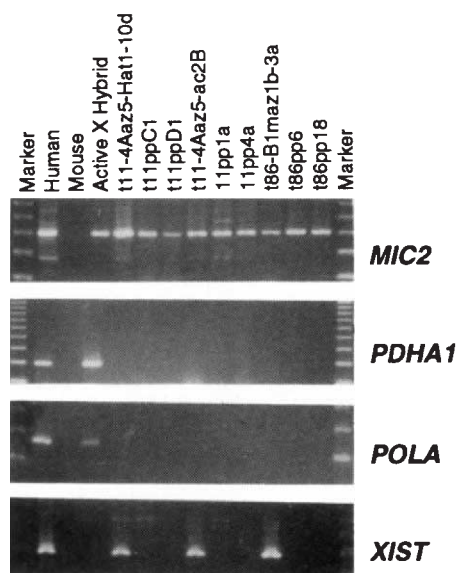


FIG. 2 RT-PCR analysis of transcription in somatic cell hybrids retaining the inactive X chromosome with or without the XIC. The outside lanes are the 123 base pair (bp) DNA size marker. The cDNAs amplified are from the left: human female; tsA1S9az31b, the parental mouse line for all the somatic cell hybrids analysed; an active X-chromosome-containing hybrid; and three sets of inactive X hybrids consisting of the parental hybrid followed by two pushmi-pullyu derivative hybrids (pp in name) which have lost the XIC (including the *XIST* gene). The location of the breakpoints in these hybrids is shown in Table 1. We have previously described the use of RT-PCR to demonstrate the transcriptional inactivation of the *POLA* gene in Xp22.1-Xp21.3 (ref. 5). In addition, *PDHA1* in Xp22.1 has been shown to be subject to X inactivation²⁶ and primers were derived spanning an intron/exon boundary for RT-PCR analyses. RNA (5 µg) was reverse-transcribed as described²³ and one-tenth of the resulting cDNA was amplified with human-specific primers for *MIC2* (ref. 23); *PDHA1* (5'-CGAATGGAATGGGAACGTCTGT-3' and 5'-CGACTTCTCGTGTACGGTAACT-3'); *POLA* (ref. 5); and *XIST* (5'-CTCCAGAT-AGCTGCAACC-3' and 5'-AGCTCCTCGGACAGCTGTAA-3').

in situ hybridization with a human X-chromosome-specific probe which showed that the X chromosome in many pushmi-pullyu hybrids was not associated with mouse chromosomal material (data not shown).

X-chromosome inactivation occurs early in development, and the resulting inactive X chromosome acquires a number of features that distinguish it from the active X chromosome, including replication, methylation, condensation and acetylation differences^{6,7,13-16}. Inactivation can be maintained in the absence of many of these features, although with an elevated reactivation frequency, suggesting that the normally very stable maintenance of X-chromosome inactivation may result from the interaction of many factors. If some component of the XIC were acting in concert with other factors to maintain inactivation, then cell lines which had lost the XIC might be expected to be more sensitive to reactivation. To examine this, two clones of the t11 series of pushmi-pullyu hybrids were serially cultured for more than 5 months, and exposed to levels of the demethylating agents 5-azacytidine and 5-azadeoxycytidine which are known to increase reactivation frequencies of inactivated genes¹⁷. No reactivation was observed (data not shown). However, without a selectable marker on the inactive X chromosome, substantial increases in the reversion frequency could have escaped detection.

The pushmi-pullyu counter-selection technique permits the selection of breakpoints between two selectable markers, in this case *UBE1* in Xp11.23 and *HPRT* in Xq26. In contrast to the active X-chromosome-containing pushmi-pullyu hybrids¹⁸ (in which the breakpoints were distributed along the length of Xq), the breakpoints isolated from the inactive X chromosome are all clustered in the proximal long arm, proximal to the XIC region (Table 1). This difference between active and inactive X pushmi-pullyu hybrids may reflect chromatin differences between active and inactive X chromosomes with respect to accessibility to breakage. However, further experiments involving isogenic active and inactive X-chromosome hybrids would be required to identify the source of this variation definitively.

XIST expression is completely concordant with inactivation in both germ cells and somatic tissues^{19,22}, indicating an early involvement in the X-inactivation process. The constitutive expression of *XIST* is also consistent with a role for the XIC in the maintenance of X-chromosome inactivation. However, based on our analysis of eight inactive X-linked genes, we conclude that the XIC is not strictly required for maintenance of transcriptional inactivity on the inactive X chromosome in somatic cells. □

Received 12 November; accepted 30 December 1993.

1. Lyon, M. F. *Nature* **190**, 372-373 (1961).
2. Russell, L. B. *Science* **140**, 976-978 (1963).
3. Lyon, M. F. in *2nd Int. Conf. congenit. Malform.* 67-68 (Int. Med. Cong., New York, 1963).
4. Mattei, M. G., Mattei, J. F., Vidal, I. & Giraud, F. *Hum. Genet.* **56**, 401-408 (1981).
5. Brown, C. J. et al. *Nature* **349**, 82-84 (1991).
6. Brown, C. J. & Willard, H. F. *Adv. dev. Biol.* **2**, 37-72 (1993).
7. Lyon, M. F. *Rev. Genet.* **26**, 17-28 (1992).
8. Brown, C. J. et al. *Nature* **349**, 38-44 (1991).
9. Brown, C. J. et al. *Cell* **71**, 527-542 (1992).
10. Brockdorff, N. et al. *Cell* **71**, 515-526 (1992).
11. Lafreniere, R. G. et al. *Genomics* **11**, 352-363 (1991).
12. Goodfellow, P., Pym, B., Mohandas, T. & Shapiro, L. J. *Am. J. hum. Genet.* **36**, 777-782 (1984).
13. Gartler, S. M., Dyer, K. A. & Goldman, M. A. in *Molecular Genetic Medicine* (ed. Friedmann, T.) 121-160 (Academic, New York, 1992).
14. Gartler, S. M. & Riggs, A. D. *Rev. Genet.* **17**, 155-190 (1983).
15. Riggs, A. D. *Aust. J. Zool.* **37**, 419-441 (1990).
16. Jeppesen, P. & Turner, B. M. *Cell* **74**, 281-289 (1993).

17. Mohandas, T., Sparkes, R. S. & Shapiro, L. J. *Science* **211**, 393-396 (1981).
18. Brown, C. J., Vivona, A., Parikh, S., Bedford, M. & Willard, H. F. *Cytogenet. Cell Genet.* **51**, 970 (1989).
19. Richler, C., Soreq, H. & Wahrman, J. *Nature Genet.* **2**, 192-195 (1992).
20. Salido, E. C., Yen, P. H., Mohandas, T. K. & Shapiro, L. J. *Nature Genet.* **2**, 196-199 (1992).
21. McCarrey, J. R. & Dilworth, D. D. *Nature Genet.* **2**, 200-203 (1992).
22. Kay, G. F. et al. *Cell* **72**, 171-182 (1993).
23. Brown, C. J., Flenniken, A. M., Williams, B. R. G. & Willard, H. F. *Nucleic Acids Res.* **18**, 4191-4195 (1990).
24. Brown, C. J. & Willard, H. F. *Am. J. hum. Genet.* **45**, 592-598 (1989).
25. Giles, R. E. & Ruddle, F. H. *Genetics* **93**, 975-996 (1979).
26. Brown, R. M., Dahl, H. H. M. & Brown, G. K. *Genomics* **4**, 174-181 (1989).

ACKNOWLEDGEMENTS. This research was supported by the NIH. We thank D. Ledbetter and L. Carrel for discussions; V. E. Powers and K. Gustashaw for cytogenetic analyses; L. Carrel for designing the *PDHA1*, *ARAF*, *ELK1* and *TFE3* gene primers; and D. Wolff for designing the *AR* gene primers.

Membrane association of Rab5 mediated by GDP-dissociation inhibitor and accompanied by GDP/GTP exchange

Oliver Ullrich, Hisanori Horiuchi, Cecilia Bucci* & Marino Zerial†

European Molecular Biology Laboratory, Postfach 10.2209, Meyerhofstrasse 1, 69012 Heidelberg, Germany

THE Rab GTPases function as specific regulators of membrane transport¹⁻⁴. The GTP/GDP cycle is believed to control shuttling of Rab proteins between the cytosol and organelle membranes. *In vitro*, Rab proteins are removed from membranes by a protein that inhibits GDP dissociation (rabGDI)⁵, which leads to formation of a cytosolic complex of Rab with the inhibitor protein⁶⁻⁸. Here we use a purified Rab5-rabGDI complex in a permeabilized cell system to investigate how the cytosolic complexed form of Rab reassociates with the membrane. We find that exogenous Rab5 is correctly targeted and induces the formation of enlarged early endosomes, demonstrating that it is functionally active. Binding of Rab5 to the acceptor membrane is accompanied by release of the rabGDI protein into the cytosol. A transient GDP-Rab5 intermediate was detected which was subsequently converted into the GTP-bound form. Our results indicate that there is a multistep mechanism for the insertion of Rab5 into the membrane which is mediated by a guanine-nucleotide-exchange factor.

Membrane-bound Rab5 was purified, pre-loaded with [³H]GDP, and incubated with purified rabGDI. On gel filtration chromatography, [³H]GDP-Rab5 co-eluted with rabGDI, indicating that the two proteins formed a complex (Fig. 1). This step was used to remove uncomplexed rabGDI, unbound [³H]GDP and detergent used to purify membrane-bound Rab5. Fractions containing Rab5 and rabGDI at 1:1 (fraction 33) and 1:5 (fractions 34-36) ratios gave identical results. The concentration of rabGDI was below that required (1 µM) to extract Rab proteins from the membrane⁵.

To test the binding of the purified Rab5-rabGDI complex to the membrane *in vitro*, Madin-Darby canine kidney (MDCK) cells were permeabilized with the bacterial toxin streptolysin O⁸ and incubated with different concentrations of Rab5-rabGDI complex for 30 min. Increasing amounts of Rab5 were found to associate with permeabilized but not with intact cells, reaching saturation at 50 nM (Fig. 2a, c). Under these conditions, 34% of exogenous Rab5 associated with membranes, corresponding to 13-fold excess over the endogenous level. In contrast, there was only a minor increase in the amount of association of rabGDI with the membrane (0.7% of added protein). Binding of Rab5 was temperature-dependent⁸, did not require cytosol and ATP, and was inhibited by 50% in the presence of 1 mM GDP (results not shown). In the absence of rabGDI or detergent, Rab5 aggregated and so failed to bind to the membrane, indicating that rabGDI is required to chaperone Rab5 during the membrane-binding process. When permeabilized cells were incubated in the presence of 50 nM Rab5-rabGDI complex for different periods of time, membrane association of Rab5 increased in a time-dependent manner, reaching a plateau after 30 min (Fig. 2b, d). A decrease in membrane-bound Rab5 was observed only after 90 min (Fig. 3d). In contrast, there was no increase in membrane-bound rabGDI with time (Fig. 2b, d).

Current models propose that Rab proteins bind to the membrane in the GTP-bound form through a process mediated by a guanine-nucleotide-exchange factor^{1-3,9}. We therefore determined the nature of the guanine-nucleotide-bound form of Rab5 that is associated with the membrane. [³H]GDP-Rab5-rabGDI complex was incubated with unpermeabilized and permeabilized MDCK cells and the radioactive GDP bound to Rab5 in the soluble and membrane fractions was measured. Dissociation of [³H]GDP from Rab5 was low when the complex was incubated with unpermeabilized cells (Fig. 3a). In contrast, in permeabilized cells, the amount of [³H]GDP associated with Rab5 recovered from the medium decreased with time. Surprisingly, increasing amounts of Rab5 in the GDP-bound form associated with the membrane of permeabilized cells, reaching a maximum (1.3 pmol) after 15 min. At this time no stoichiometric increase in the membrane association of rabGDI could be detected (Fig. 2b, d), so we conclude that membrane-bound [³H]GDP-Rab5 was not complexed with rabGDI. Membrane-bound radiolabelled nucleotide decreased to background level after 60 min (Fig. 3b), however, and given that at this time

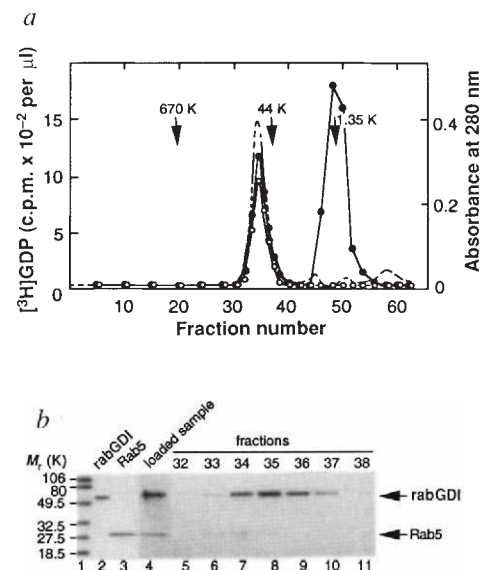


FIG. 1 Purification of [³H]GDP-Rab5-rabGDI complex by gel filtration column chromatography. a, Elution profile shows total radioactivity (●) and protein-bound radioactivity (○) as measured by filter-binding⁵. Relative molecular masses are indicated in thousands (arrows) b, Coomassie blue-stained SDS-polyacrylamide gel showing purified His₆-tagged rabGDI (lane 2), Rab5 (lane 3), and the Rab5-rabGDI complex after gel filtration chromatography (fractions 32-38, lanes 5-11). Protein-bound radioactivity (a) appeared in the same fractions (33-37) in which Rab5 and rabGDI were detected by SDS-PAGE (b). For total [³H]GDP radioactivity (a), an additional peak was detected (fractions 46-52) corresponding to free [³H]GDP.

METHODS. Rab5 was expressed in *Spodoptera frugiperda* Sf9 cells using pVL1393 baculovirus transfer vector. The purification of Rab5 will be described in detail elsewhere (H.H. et al., manuscript in preparation). Briefly, Rab5 was extracted from the membrane of infected Sf9 cells using 0.6% CHAPS and purified on a Mono Q HR 5/5 column (Pharmacia). rabGDI cDNA¹⁵ was cloned into plasmid pRSET A (Invitrogen) and His₆-tagged rabGDI expressed in *E. coli* was purified on Ni²⁺-NTA-agarose (Qiagen) and Mono Q HR 5/5 columns. The [³H]GDP-Rab5-rabGDI complex was obtained by first incubating 1.5 µM Rab5 with 3 µM [³H]GDP as described⁵, followed by incubation with 3 µM rabGDI and 20 mM MgCl₂ at 30 °C for 10 min. The sample (0.5 ml) was then loaded onto a Superose-12 gel filtration column (Pharmacia) equilibrated with buffer A (50 mM HEPES/KOH, pH 7.1, 4 mM MgCl₂, 10 mM EGTA, 8.37 mM CaCl₂, 78 mM KCl, 1 mM DTT) containing 1 µM GDP and 0.005% Triton X-100. In the absence of rabGDI, no Rab5 was recovered (not shown). The stoichiometry of Rab5 and rabGDI in the fractions was quantified by western blot analysis using purified Rab5 and rabGDI as standards^{8,11}.

* Present address: Dipartimento di Biologia e Patologia Cellulare e Molecolare 'L. Califano', Via S. Pansini 5, 80131 Napoli, Italy.

† To whom correspondence should be addressed.

2.3 pmol Rab5 was membrane-bound (as measured by western blotting; Fig. 3d), these results indicate that [3 H]GDP-Rab5 associates with the membrane but that this process is followed by release of GDP. To show that GTP is exchanged for GDP bound to Rab5, cells were incubated with GDP-Rab5-rabGDI complex in the presence of [35 S]GTP γ S for different times. Membrane-associated Rab5 was quantitatively immunoprecipitated and the amount of radioactivity bound to Rab5 determined. Figure 3c shows that in permeabilized but not in unpermeabilized cells there was a time-dependent increase in Rab5-bound [35 S]GTP γ S. After 60 min incubation, 2.4 pmol [35 S]GTP γ S-bound Rab5 was recovered from the membrane. The initial lag period could be due to slow diffusion of the Rab5-rabGDI complex into the cells. At any time, the sum of radiolabelled [3 H]GDP-Rab5 and [35 S]GTP γ S-Rab5 closely matched the amount of membrane-bound Rab5 estimated from western blot analysis (Fig. 3d).

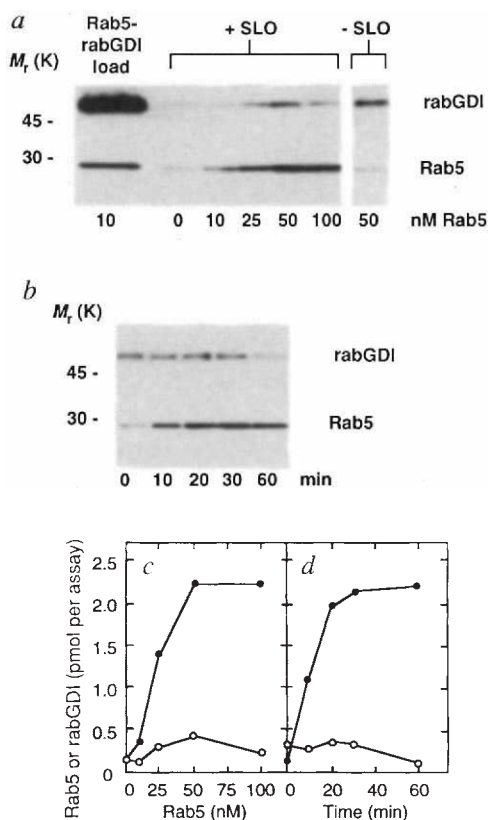


FIG. 2 Dose- (a, c) and time- (b, d) dependent saturation of Rab5 binding to the membrane of streptolysin O (SLO)-permeabilized cells. Western blot analysis (a, b) and quantification (c, d) of Rab5 (●) and rabGDI (○) binding to permeabilized cells. Permeabilized and unpermeabilized cells were incubated (a, c) for 30 min with various concentrations (0–100 nM) of purified Rab5–rabGDI complex, or (b, d) with 50 nM Rab5–rabGDI complex for different times. Binding of Rab5 and rabGDI to cells was quantified by western blot analysis.

METHODS. MDCK cells were grown to confluence in 24-well plastic dishes (3×10^6 cells) and permeabilized with SLO as described⁵. Cells were placed on ice for 30 min with two changes of buffer A (Fig. 1 legend) to deplete cytosol. Incubation of cells with Rab5–rabGDI complex was at 37 °C in the presence of 50 μ M GTP γ S. GTP γ S was used as for Fig. 3 and similar results were obtained with GTP. Cells were then placed on ice, incubation buffer removed, and the scraped cells washed for 20 min in buffer A. Cells were boiled in SDS sample buffer and analysed by SDS-PAGE and western blotting using monoclonal anti-Rab5 (C.B. et al., manuscript submitted) and polyclonal anti-rabGDI⁸ antibodies. Bound antibodies were visualized by the enhanced chemiluminescence detection system (Amersham Corp.) and quantified by densitometry (Molecular Dynamics).

The presence of membrane-bound GDP-Rab5 may reflect a rapid equilibrium between soluble GDP-Rab5–rabGDI complex and membrane-associated Rab5, rather than a functional intermediate in the formation of the stable membrane-bound GTP-bound form. To distinguish between these possibilities, GDP-Rab5–rabGDI was preincubated with permeabilized cells

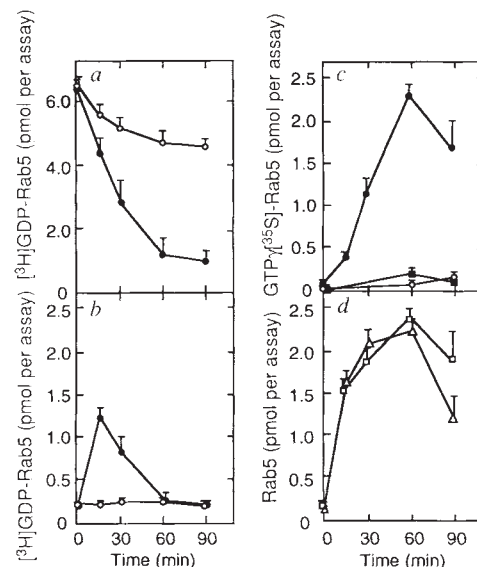


FIG. 3 a and b, Time course of [3 H]GDP-Rab5 binding to the membrane of SLO-permeabilized cells. [3 H]GDP-Rab5–rabGDI complex was incubated with SLO-permeabilized and unpermeabilized cells for different times. The amount of [3 H]GDP-Rab5 recovered from medium (a) and cells (b) was measured. c, Time course of [35 S]GTP γ S binding to Rab5 on the membrane of permeabilized cells. Unlabelled Rab5–rabGDI complex was incubated with cells in the presence of [35 S]GTP γ S for the indicated times. Membrane-bound [35 S]GTP γ S-Rab5 was measured after immunoprecipitation. In d, the sum of membrane-bound [3 H]GDP-Rab5 and [35 S]GTP γ S-Rab5 is compared with the amount of membrane-bound Rab5 as quantified by western blotting. Symbols used in a–c: (●) permeabilized cells with Rab5–rabGDI complex; (○) unpermeabilized cells with Rab5–rabGDI complex; (■) permeabilized cells without Rab5–rabGDI complex; in d: (□) sum of membrane-bound [3 H]GDP-Rab5 (from b) and [35 S]GTP γ S-Rab5 (from c); (△) membrane-associated Rab5 (from b and c) quantified by western blotting.

METHODS. a and b, 6.5 pmol [3 H]GDP-Rab5–rabGDI complex (54 nM, 0.8×10^4 c.p.m. pmol^{-1}) were incubated in the presence of 50 μ M GTP γ S with cells (Fig. 2 legend). After the indicated times, medium was collected and adjusted to 220 μ l, to contain 1% NP-40, 20 mM MgCl₂, and 1 mM GTP. Cells were scraped and washed for 10 min in buffer B (50 mM Tris-HCl, pH 7.5, 20 mM MgCl₂, 100 mM NaCl, 1 mM GTP), followed by solubilization in 220 μ l buffer B containing 1% NP-40. Aliquots of media and solubilized cell fractions were assessed for protein-bound [3 H]GDP by filter binding⁵. This method was used because most [3 H]GDP was associated with Rab5, and in addition, identical results were obtained by immunoprecipitation. In c, 6.5 pmol unlabelled Rab5–rabGDI complex were incubated in the presence of 50 μ M [35 S]GTP γ S (1.0×10^4 c.p.m. pmol^{-1}) with the cells. The non-hydrolysable analogue was used instead of GTP to study GDP/GTP exchange independent of GTP hydrolysis. After incubation, cells were washed and solubilized as already described. Rab5 was immunoprecipitated from the cell lysates using monoclonal anti-Rab5 antibodies (Fig. 2 legend). Immunoprecipitation of Rab5 was necessary owing to significant binding of [35 S]GTP γ S to other cellular proteins. The amount of [35 S]GTP γ S-Rab5 associated with anti-Rab5 antibodies bound to protein A-agarose was determined after subtraction of the radioactivity obtained with protein A-agarose alone. Efficiency of immunoprecipitation was >75%, as determined by immunoprecipitating Rab5 preloaded with [35 S]GTP γ S added to cell lysates. In d, aliquots of cell lysates were analysed by quantitative western blotting (Fig. 2 legend) using purified Rab5 as standard. Bars represent standard error of the mean of four independent experiments.

for 15 min, unbound complex removed, and the cells were then incubated for 15 or 30 min in the presence of [35 S]GTP γ S. At these times, 27 and 50% of membrane-associated Rab5 was detected in the [35 S]GTP γ S-bound form, respectively. These results show that GDP-Rab5 is converted into the GTP-bound form following membrane binding.

We next investigated whether Rab5 was properly localized and functionally active using confocal immunofluorescence microscopy. Because of the availability of antibodies, we used A431 cells permeabilized with streptolysin O to determine whether exogenous Rab5 co-localizes with human transferrin receptor-containing structures, as demonstrated earlier^{10,11}. Binding of Rab5 to A431 cells and to MDCK cells was similar (not shown). Exogenous Rab5 extensively co-localized with human transferrin receptor on the plasma membrane and cytoplasmic vesicles (Fig. 4*a-c*) but not with the late endosome/lysosome marker lamp-1 (ref. 12) (Fig. 4*d-f*) or with the endo-

plasmic reticulum (ER)-Golgi intermediate compartment/*cis*-Golgi marker p28 (V. N. Subramaniam *et al.*, manuscript submitted). These results indicate that, like overexpressed Rab5 in intact cells^{10,11}, purified Rab5 was correctly targeted in permeabilized cells. Incubation of cells in the presence of the Rab5-rabGDI complex for 5, 15 and 30 min resulted in a time-dependent increase in the size of early endosomes (compare Fig. 4*a, d* with *g, h*), as previously shown for cells transiently overexpressing Rab 5 (ref. 11). In contrast, there was no increase in the size of early endosomes when incubation was in the presence of 1 mM GDP to reduce GDP/GTP exchange (compare Fig. 4*g* and *i*). As under these conditions membrane association of Rab5 is inhibited but not abolished (see above), we conclude that the activity of Rab5 in expanding the size of early endosomes requires GTP.

The existence of Rab proteins in both cytosolic and membrane-bound pools is thought to reflect their shuttling

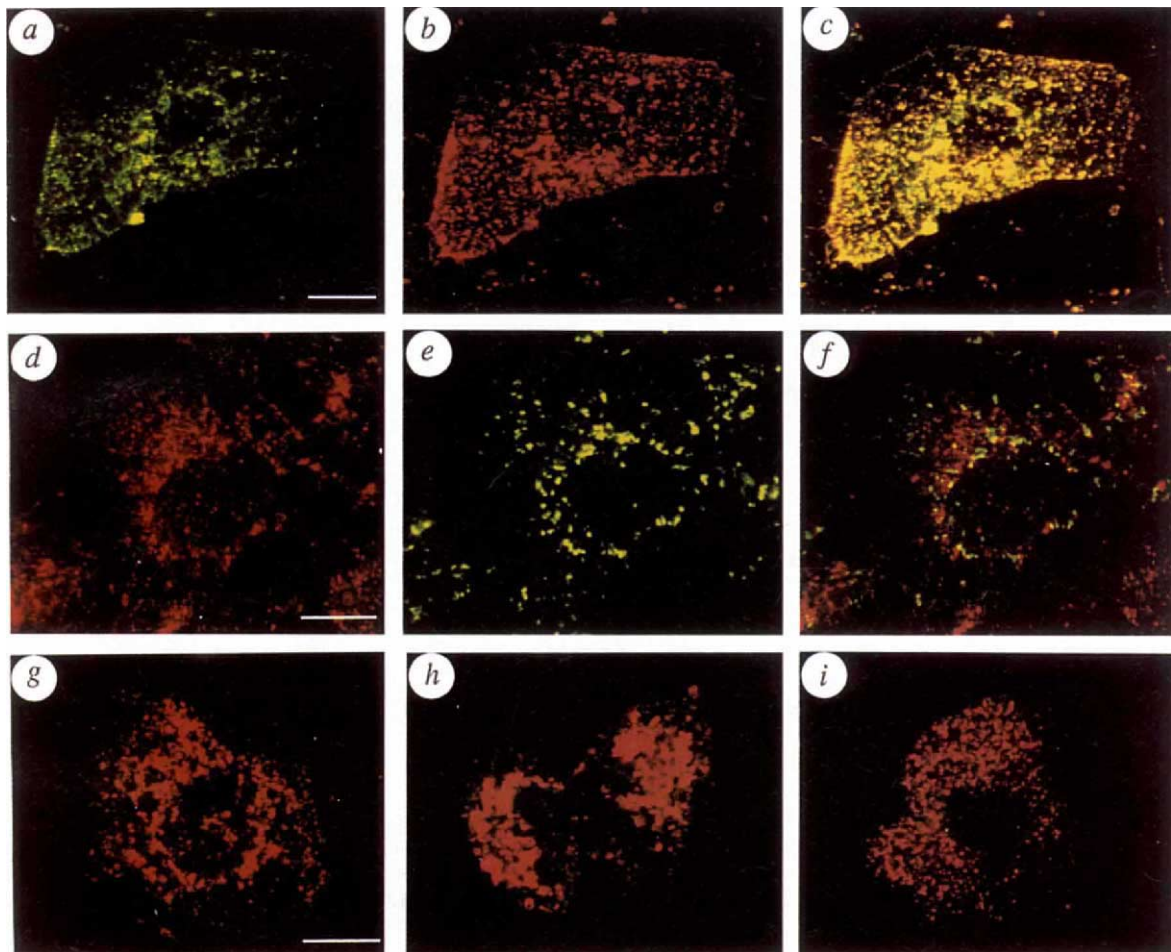


FIG. 4 Confocal fluorescence microscopy localization of Rab5, human transferrin receptor (hTR) and human (h)lamp-1 in SLO-permeabilized A431 cells incubated with Rab5-rabGDI for 5 min (*a-f*), 15 min (*g, i*) and 30 min (*h*). Incubations were done in the presence of 100 μ M GTP (*a-h*) and 1 mM GDP (*i*); *a, d, g, h* and *i*, show Rab5; *b, hTR*; *c*, superimposition of images in *a* and *b* (Rab5, hTR); *e, h-lamp-1*; *f*, superimposition of images in *d* and *e* (Rab5, h-lamp-1). Yellow indicates areas of co-localization. Co-localization of Rab5 and hTR was seen for all incubation times (data not shown). Exogenous Rab5 was detected using antibody concentrations insufficient to detect the endogenous protein. Addition of Rab5-rabGDI complex to unpermeabilized cells or buffer containing 100 μ M GTP alone to permeabilized cells did not affect the morphology of hTR-positive structures (data not shown).

METHODS. A431 cells grown on glass coverslips were permeabilized for 5 min with 5-fold less SLO than was used for MDCK cells⁸ and

immediately incubated without washing, with Rab5-rabGDI complex at 37 °C. Under these conditions, cells were only partially depleted of cytosol and ATP, as demonstrated by quantification of lactate dehydrogenase release¹⁶. Medium was then removed and cells permeabilized with 0.05% saponin before fixation to wash out unbound Rab5-rabGDI complex. Double immunofluorescence staining was carried out as described¹⁷. As first antibodies we used affinity-purified rabbit anti-Rab5 polyclonal antibodies (*a*), mouse anti-Rab5 monoclonal antibodies (*d, g, h, i*), mouse anti-hTR monoclonal antibodies (Boehringer) (*b*), and rabbit anti-human lamp-1¹² polyclonal antibodies (*e*). Secondary antibodies were FITC-labelled donkey anti-rabbit antibodies and rhodamine-labelled donkey anti-mouse antibodies. Coverslips were viewed with the confocal microscope developed at EMBL. Excitation wavelengths were 476 nm and 529 nm; sections were 0.5 μ m. Scale, 10 μ m.

between donor and acceptor compartments to serve multiple rounds of vesicular transport¹⁻⁴. Our *in vitro* results suggest a mechanism by which the membrane association of Rab5 could be regulated. First, this process requires rabGDI which is released upon binding of Rab5 to the membrane. Second, we find that Rab5 binds to the membrane in the GDP-bound form initially, despite the release of rabGDI. This is consistent with the finding that the membrane association of a Rab5 mutant that has a higher affinity for GDP than for GTP is not impaired¹³. Third, the GDP-bound form is then converted into the GTP-bound form. Thus, membrane association of Rab5 appears to be a multistep process involving one or more factors which bind GDP-Rab5, induce dissociation from rabGDI, and catalyse the conversion from the GDP-bound to the GTP-bound form. The excess of Rab5-rabGDI complex might saturate this factor(s) and allow the detection of the membrane-bound GDP-Rab5 intermediate which would otherwise be difficult to detect under physiological conditions. High concentrations of GDP inhibit membrane association and prevent the formation of large endocytic structures induced by Rab5, indicating that GDP/GTP exchange may be rate-limiting for membrane binding and necessary for the function of Rab5, although we cannot exclude the participation of other GTP-binding proteins in this process.

A mammalian exchange factor for Rab proteins, Mss4, has been shown to stimulate GDP release from Sec4p, Rab3A and Ypt1p¹⁴. As membrane localization of Rab proteins is a specific

process¹⁰, membrane association of Rab5 may be mediated by a specific exchange factor that is distinct from Mss4 but perhaps related to it. □

Received 12 October 1993; accepted 21 January 1994.

1. Bourne, H. R. *Cell* **53**, 669-671 (1988).
2. Takai, Y., Kaibuchi, K., Kikuchi, A. & Kawata, M. *Int. Rev. Cytol.* **133**, 187-231 (1992).
3. Pfeffer, S. R. *Trends Cell Biol.* **2**, 41-46 (1992).
4. Zerial, M. & Stenmark, H. *Curr. Opin. Cell Biol.* **5**, 613-620 (1993).
5. Sasaki, T. *et al. J. biol. Chem.* **265**, 2333-2337 (1990).
6. Araki, S., Kikuchi, A., Hata, Y., Isomura, M. & Takai, Y. *J. biol. Chem.* **265**, 13007-13015 (1990).
7. Soldati, T., Riederer, M. A. & Pfeffer, S. R. *Molec. Biol. Cell* **4**, 425-434 (1993).
8. Ullrich, O. *et al. J. biol. Chem.* **268**, 18143-18150 (1993).
9. Goud, B. & McCaffrey, M. *Curr. Opin. Cell Biol.* **3**, 626-633 (1991).
10. Chavrier, P. *et al. Nature* **353**, 769-772 (1991).
11. Bucci, C. *et al. Cell* **70**, 715-728 (1992).
12. Carlsson, S. R., Roth, J., Piiler, F. & Fukuda, M. *J. biol. Chem.* **263**, 18911-18919 (1988).
13. Stenmark, H. *et al. EMBO J.* (in the press).
14. Burton, J., Roberts, D., Montaldi, M., Novick, P. & Camilli, P. D. *Nature* **361**, 464-467 (1993).
15. Matsui, Y. *et al. Molec. cell. Biol.* **10**, 4116-4122 (1990).
16. Kobayashi, T., Pimplikar, S. W., Parton, R. G., Bhakdi, S. & Simons, K. *FEBS Lett.* **300**, 227-231 (1992).
17. Chavrier, P., Parton, R. G., Hauri, H. P., Simons, K. & Zerial, M. *Cell* **62**, 317-329 (1990).

ACKNOWLEDGEMENTS. We thank S. Carlsson and M. Fukuda for antibody against lamp-1, and W. Hong for antibody against p28. M. Stapleton and A. Giner for technical assistance; A. Sossick for photography; and B. Hoflack, C. Murphy, K. Simons and H. Stenmark for discussion and for critically reading the manuscript. O.U. holds a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft and H.H. holds an EMBO long-term fellowship. This study was supported by a grant from the Human Frontier Science Program, and by the Japan Foundation for Ageing and Health (H.H.).

Regulation of TFIIF ATPase and kinase activities by TFIIE during active initiation complex formation

Yoshiaki Ohkuma & Robert G. Roeder

Laboratory of Biochemistry and Molecular Biology, Rockefeller University, New York, New York 10021, USA

THE general transcription factor TFIIE, together with other general transcription factors, is essential for transcription initiation by RNA polymerase II¹⁻⁵. TFIIE stimulates the TFIIF-dependent kinase activity that phosphorylates the carboxy-terminal domain of the largest subunit of RNA polymerase II⁶, and possesses a helicase activity⁷. Here we show that human TFIIF has DNA-dependent ATPase activity and we characterize the stimulatory effect of TFIIE on both the ATPase and kinase activities. We demonstrate that extensive phosphorylation of RNA polymerase II occurs in a TFIIE-dependent manner in both the absence and presence of DNA but, in the latter case, only at a late stage of preinitiation complex assembly. We also show that TFIIF specifically phosphorylates three general transcription factors, human TFIID⁸ (TBP), TFIIE- α and TFIIF- α (RAP74).

To characterize the function of TFIIE in transcription initiation, we have set up a reconstituted system using highly purified factors. TFIIF was separated from TFIIF by further purification and, as shown in Fig. 1a, contained seven polypeptides (M_r s 94K, 75K, 62K, 46K, 43K, 38K and 35K) similar to those reported for the highly purified human BTF2 (TFIIF)⁸ and the rat homologue δ (ref. 9). The 62K polypeptide (p62) reacts with a monoclonal antibody (3c9) directed against it, and was isolated as one of the integral TFIIF (BTF2) subunits (Fig. 1b)⁸. These TFIIF fractions support basal transcription (Fig. 1c).

Hydrolysis of the β - γ bond of ATP (dATP) is essential for transcription initiation^{10,11}, and part of the energy might be used

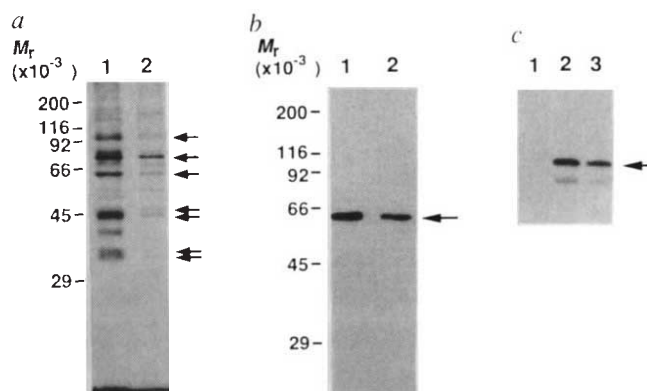


FIG. 1 The general transcription factor TFIIF. *a*, SDS-PAGE of TFIIF fractions. HPLC-heparin 5PW fractions of TFIIF were applied to a 10% polyacrylamide gel and the gel was stained with silver (Bio-Rad). The same volume (10 μ l) of fractions 50 (lane 1) and 52 (lane 2) was used. Arrows indicate putative TFIIF subunits. *b*, Immunoblotting with anti-p62 monoclonal antibody. After SDS-PAGE of the same volume of fractions 50 (lane 1) and 52 (lane 2), proteins were blotted to a nylon membrane (Immobilon-P; Millipore). Arrow indicates p62 polypeptide. *c*, Complementation of basal transcription activity with TFIIF fractions. Reactions containing the activity with TFIIF fractions. Reactions containing the adenovirus-2 major late promoter template, pML(C₂)A-50 (ref. 25), RNA polymerase II and all general transcription factors except TFIIF (refs 25, 26) were complemented with buffer only (lane 1) and 1 μ l of fraction 50 (lane 2) or 52 (lane 3). Arrow indicates the position of the specific transcript.

METHODS. TFIIF was purified through six chromatographic steps (phosphocellulose, DEAE-cellulose, Sephacryl S-300, HPLC-DEAE, HPLC-SP and HPLC-heparin), with gradient elution on the last three columns. TFIIF was eluted at 370 mM KCl from HPLC-heparin and completely separated from TFIIF. A monoclonal antibody against p62 (3c9) was used as described⁷. Transcription assays were performed as described^{1,25}, except the HPLC-DEAE fraction of TFIIF/TFIIF were replaced by the HPLC-heparin fractions of TFIIF and TFIIF.

for unwinding the DNA double helix around the initiation site. The finding of DNA-dependent ATPase activity in the rat (factor δ)⁹ and yeast (factor b)¹² homologues suggested that human TFIH would also have such activity, and this was confirmed (Fig. 2; compare lanes 5 and 6 with lanes 1 and 2). Most interestingly, especially in light of the effect of TFIIE on a second TFIH activity (see below), TFIIE stimulated this ATPase activity about fivefold (Fig. 2; compare lanes 9 and 10 with lanes 5 and 6 and 7 and 8).

In addition to an ATPase activity, TFIH and its yeast and rat homologues possessed a kinase activity that phosphorylates the heptapeptide(YSPSPS)¹³-repeat-containing C-terminal domain (CTD) in the large subunit of mammalian RNA polymerase II (refs 6, 12, 14). The CTD interacted with both TFIID and the derived TATA-box-binding protein TBP (TFIID τ) alone^{15,16} and based on sequence similarity with the SPKK motif in histone H1 (ref. 17), was predicted to interact with DNA. The highly purified preparation of TFIH used here also contained a CTD kinase activity, which co-eluted with the ATPase activity and with TFIH transcription activity through six chromatographic steps (Fig. 1 legend). As further evidence of the integral association of these activities with TFIH, the anti-p62 monoclonal antibody 3c9, previously shown to inhibit TFIH activity⁶, co-immunoprecipitated the ATPase and CTD kinase activities along with p62 (data not shown).

We next examined the effect of TFIIE on the kinase activity in the presence and absence of DNA (Fig. 3a). Without DNA, tetrameric TFIIE markedly stimulated CTD phosphorylation (in intact RNA polymerase II) and this depended entirely on the

presence of TFIH (Fig. 3a; compare lanes 9 with lanes 6 and 2). Somewhat surprisingly, the isolated α - and β -subunits of TFIIE also produced a substantial CTD phosphorylation in the presence of TFIH (Fig. 3a; compare lanes 7 and 8 with lane 6). Without DNA, other general factors (TBP, TFIIB and TFIIF) necessary for preinitiation complex formation had no significant effect on TFIIE-dependent phosphorylation of the CTD by TFIH (Fig. 3a; compare lanes 14, 18 and 22 with lane 9), nor did they stimulate CTD phosphorylation by TFIH in the absence of TFIIE (compare lanes 12, 16 and 20 with lane 6). In marked contrast, in the presence of DNA, optimal phosphorylation of the CTD required all the general factors and RNA polymerase II (Fig. 3a, lanes 13, 15, 17, 19, 21 and 23). Lu *et al.*⁶ also reported CTD phosphorylation in the presence of DNA and general factors, but this occurred as a gradual stepwise increase during the sequential addition of individual general factors. Their results contrast with the present results which show clearly that TFIH cannot phosphorylate the CTD efficiently without TFIIE, and that there is a large (10-fold) stimulation by TFIIE which requires all other general initiation factors. Our results are more satisfying in showing activation of TFIH at a specific time in preinitiation complex assembly, and by the factor (TFIIE) directly required for its recruitment. The basis for the discrepancy is unclear, but may reflect the use of a heterologous TBP by Lu *et al.*⁶ and possible cross-contamination of general transcription factors with TFIIE.

Although previous studies have implicated only the CTD as a major substrate for the TFIH-associated kinase, some general factors also seem to be substrates. TFIIE- α was phosphorylated in a TFIH-dependent manner (Fig. 3b; compare lanes 7 and 1) and this phosphorylation was enhanced by TFIIE- β (compare lanes 5 and 7). TFIH also phosphorylated human TBP, and TFIIE- α stimulated this reaction (Fig. 3b, lanes 8–11; Fig. 3c, lanes 8–10). Phosphorylation of TBP was specific for the amino-terminal region (residues 1–154) of human TBP, as the carboxy-terminal core domain (residues 155–335) (Fig. 3c, lanes 11–13) and yeast TBP (lanes 14–16) were not phosphorylated by TFIH. The CTD-like SPT (serine, proline and threonine-rich) sequences in the amino terminus of TBP, which is highly conserved in higher eukaryotes¹⁸, could be the targets for phosphorylation. TFIH also phosphorylated TFIIF- α (RAP74) independently of TFIIF- β (RAP30) (Fig. 3c, lanes 17, 20 and 23) or TFIIE (Fig. 3c, lanes 17–19, 23 and 24). TFIIB and TFIIF- β were not phosphorylated by the TFIH kinase in the absence or presence of TFIIE (Fig. 3b, lanes 12–14; Fig. 3c, lanes 20–22). Phosphoamino-acid analyses indicated that phosphorylation was at serine (93%) and threonine (7%) on the CTD of RNA polymerase II, only at serine on TFIIE, and at serine (34%) and threonine (66%) on human TBP (data not shown). Although the significance of general initiation factor phosphorylation is unclear, it could affect protein-protein interactions in the initiation complex and, ultimately, either basal or activator-dependent transcription initiation.

Early studies demonstrated that CTD phosphorylation is associated with transition from the initiation to the elongation phase of transcription^{19,20}, suggesting a causal role in this process. Consistent with this view, other studies have indicated that the non-phosphorylated form (IIA) of RNA polymerase II enters the complex more efficiently than does the phosphorylated form (IIO)^{21,22}. Based on studies with a kinase inhibitor that failed to inhibit basal transcription, Serizawa *et al.* have recently argued for a dominant, if not exclusive, role for CTD phosphorylation in preinitiation assembly²³. However, it is possible that phosphorylation inhibition was incomplete in this analysis. Moreover, the results do not exclude a role for CTD phosphorylation in the initiation/elongation transition under more physiological conditions. Hence, it remains possible that CTD phosphorylation is involved in both preinitiation complex assembly (RNA polymerase entry) and subsequent initiation or elongation transitions, and thus that functional interactions

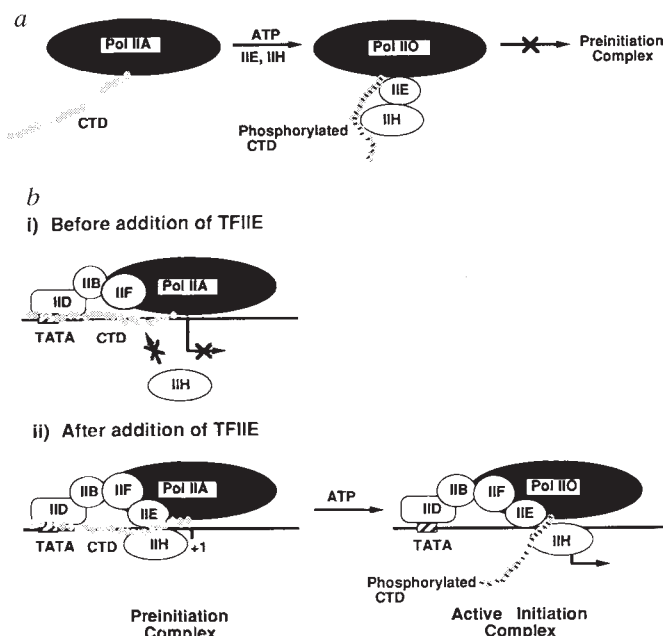


FIG. 4 Proposed roles of TFIIE and TFIH in initiation complex formation. a, In the absence of template DNA, the CTD of RNA polymerase II (IIA) is exposed in solution and can therefore be strongly phosphorylated by TFIH with the help of TFIIE. Phosphorylated RNA polymerase II (IIO) cannot efficiently enter the preinitiation complex formation pathway^{21,22}. b, In the presence of template, TFIH cannot act on the CTD because of its tight association with both DNA and TFIID (or TBP) in the absence of TFIIE. Addition of TFIIE enables TFIH to enter the preinitiation complex²⁹ either by a tethering interaction with TFIIE or by a conformational change in the preinitiation complex, or both. TFIH phosphorylates the CTD and causes a further conformational change of the preinitiation complex which, together with melting of the DNA duplex around the initiation site using energy supplied from ATP hydrolysis, leads to formation of an active initiation complex.

between TFIIE and TFIIH are important for these events.

Thus in the absence of template, TFIIE and TFIIH could cooperate to phosphorylate the CTD and prevent further progression towards formation of a non-productive complex (Fig. 4). In the presence of template, TFIIH might join the preinitiation complex and act on the CTD but only at a late stage in the presence of TFIIE. TFIIE could induce conformational changes in the preinitiation complex that trigger CTD phosphorylation by TFIIH and, consequently, conversion into an active initiation complex. TFIIE/TFIIH-dependent CTD phosphorylation might also be involved in early elongation, as RNA polymerase II molecules that pause in the 5' regions of *Drosophila hsp70* and *hsp26* genes are not phosphorylated (T. O'Brian and J. Lis, personal communication). The ATPase activity of

TFIIH might contribute to phosphate use in CTD phosphorylation, but it might also have a separate function (for example, in DNA melting²⁴) because ATP is required even when the RNA polymerase II (IIB) lacking the CTD is used for transcription²⁰. The intrinsic kinase and ATPase activities in TFIIH may indeed be differentially regulated, because the latter activity can be stimulated by various single- and double-strand DNA molecules⁹, whereas the former requires DNA elements containing the TATA box and/or the initiator sequence to form a complete preinitiation complex (refs 6, 14 and data not shown). Given the pivotal role of TFIIE in controlling these TFIIH-mediated steps in the initiation pathway, structure-function analyses of TFIIE should help to elucidate the associated regulatory mechanism(s). □

Received 15 September 1993; accepted 25 January 1994.

- Ohkuma, Y. et al. *Nature* **354**, 398–401 (1991).
- Sumimoto, H. et al. *Nature* **354**, 401–404 (1991).
- Roeder, R. G. *Trends Biol. Sci.* **16**, 402–408 (1991).
- Zawal, L. & Reinberg, D. *Prog. Nucleic Acid Res. molec. Biol.* **44**, 67–108 (1993).
- Inostroza, J., Flores, O. & Reinberg, D. *J. Biol. Chem.* **266**, 9304–9308 (1991).
- Lu, H., Zawal, L., Fischer, L., Egly, J.-M. & Reinberg, D. *Nature* **358**, 641–645 (1992).
- Schaeffer, L. et al. *Science* **260**, 58–63 (1993).
- Fischer, L. et al. *Science* **257**, 1392–1395 (1992).
- Conaway, R. C. & Conaway, J. W. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7356–7360 (1989).
- Sawadogo, M. & Roeder, R. G. *J. Biol. Chem.* **259**, 5321–5326 (1984).
- Rappaport, J. & Weinmann, R. *J. Biol. Chem.* **262**, 17510–17515 (1987).
- Feaver, W. J., Gileadi, O., Li, Y. & Kornberg, R. D. *Cell* **67**, 1223–1230 (1991).
- Corden, J. L., Cadena, D. L., Ahearn, J. H. & Dahmus, M. E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7934–7938 (1985).
- Serizawa, H., Conaway, R. C. & Conaway, J. W. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7476–7480 (1992).
- Conaway, R. C., Bradsher, J. N. & Conaway, J. W. *J. Biol. Chem.* **267**, 8464–8467 (1992).
- Usheva, A. et al. *Cell* **69**, 871–881 (1992).
- Suzuki, M. *Nature* **344**, 562–565 (1990).
- Hoffmann, A. et al. *Nature* **346**, 387–390 (1990).
- Payne, J. M., Laybourn, P. J. & Dahmus, M. E. *J. Biol. Chem.* **264**, 19621–19629 (1989).

- Laybourn, P. J. & Dahmus, M. E. *J. Biol. Chem.* **265**, 13165–13173 (1990).
- Chesnut, J. D., Stephens, J. H. & Dahmus, M. E. *J. Biol. Chem.* **267**, 10500–10506 (1992).
- Lu, H., Flores, O., Weinmann, R. & Reinberg, D. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10004–10008 (1991).
- Serizawa, H., Conaway, J. W. & Conaway, R. C. *Nature* **363**, 371–374 (1993).
- Wang, W., Carey, M. & Gralla, J. D. *Science* **255**, 450–453 (1992).
- Sawadogo, M. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4394–4398 (1985).
- Sumimoto, H., Ohkuma, Y., Yamamoto, T., Horikoshi, M. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **87**, 9158–9162 (1990).
- Ohkuma, Y., Sumimoto, H., Horikoshi, M. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **87**, 9163–9167 (1990).
- Seki, M. et al. *Biochemistry* **25**, 3239–3245 (1986).
- Flores, O., Lu, H. & Reinberg, D. *J. Biol. Chem.* **267**, 2786–2793 (1992).

ACKNOWLEDGEMENTS. We thank A. Hoffmann, D. K. Lee, T. Oelgeschlaeger and K. Wang for critical reading of the manuscript, J.-M. Egly for the monoclonal antibody (3c9) against the TFIIH p62 subunit, T. Eki and T. Enomoto for ATPase B, T. Suzuki for technical assistance and advice, and our colleagues for discussion. Y.O. was supported partly by a fellowship from the Human Frontiers Scientific Programme Organisation and partly by a biomedical fellowship from Charles H. Revson Foundation. This study was supported by NIH grants to R.G.R. and by the Pew Trusts to Rockefeller University.

Detection of localized DNA flexibility

Jason D. Kahn*, Elizabeth Yun†
& Donald M. Crothers*††

Departments of * Chemistry and

† Molecular Biophysics and Biochemistry, Yale University,
New Haven, Connecticut 06511, USA

THE bending and flexibility of DNA are important in packaging, recombination and transcription^{1–3}. Bending decreases electrophoretic mobility in a manner depending on bend position within a fragment (circular permutation⁴) and on the distance between bends (phasing analysis^{5,6}). Bending can also affect DNA ring closure (cyclization^{7–10}). The lack of a complete theory for the mechanism of gel retardation hampers measurement of bend magnitudes by electrophoresis, whereas cyclization is done entirely in solution and is well understood theoretically⁹. Disagreements between bend angles estimated by the two electrophoretic assays have been ascribed to DNA flexibility¹¹. Here we test this interpretation using an internal loop as a model flexible locus. Whereas the circular permutation and helical phasing experiments are only subtly affected by the loop, DNA cyclization kinetics detects and quantifies substantial increases in torsional and bending flexibility. Furthermore, the results support a functional role for the stress of DNA bending in inducing base-pair opening¹².

Electrophoretic mobility is maximally decreased by static bending at the centre of a DNA fragment, a phenomenon rationalized by a reduction in end-to-end distance^{4,13}. But simple reptation theory for electrophoresis does not predict bending-induced retardation¹⁴, which probably results from interactions between

bends and the gel matrix which trap DNA temporarily^{14,15}. Accordingly, a flexible site might not affect mobility. We prepared molecules with six-nucleotide internal loops (three consecutive mismatches; Fig. 1), tested the loops' effects on electrophoresis, and compared the results with those from cyclization kinetics.

Flexible loops reduce gel mobility in the circular permutation experiment by less than 5% (Fig. 2). Except for fraying of loops near the end¹⁶, retardation is maximal for a centrally positioned loop. (Fraying may affect the mobility of nicked DNAs, which are more retarded than looped DNA, in a position-dependent way¹⁷.) Pyrimidine loops reduce mobility more than purine loops, which may stack to resemble double-stranded (ds) DNA. Loop retardation, as measured by R_L (Fig. 2), is only slightly dependent on gel percentage, in contrast to the properties of curved molecules^{18,19}. A-tract-bend calibration curves (8% gel)¹⁹, with an 18° bend per A-tract²⁰, give bend angles of 26° (1.4 A-tract equivalents) for pyrimidine loops and 21° for purine loops, whereas bend angles (α) of 30° and 24° are obtained from the empirical equation $\mu_m/\mu_c = \cos(\alpha/2)$, where μ_m and μ_c are extrapolated mobilities for DNA bent at the middle and the end^{11,13}. A flexible loop in the permutation experiment acts approximately like a single A-tract, near the limit of reliable detectability.

Helical phasing analysis^{5,6}, in which the distance between two bends is varied and gel mobility typically decreases maximally for bends in *cis*, also shows very little effect due to the loop (Fig. 3). There is no dependence on the distance between a static A-tract bend and the loop, indicating that any loop-induced bending or flexibility has no preferred direction. The w_l (with loop) DNA mobility relative to all-ds DNA increases slightly with gel percentage, the opposite of the effect for additional static bending. The flexible site might be damping interactions between the static bend and the gel, thereby allowing the DNA to move more rapidly. Our results show that increased flexibility

† To whom correspondence should be addressed.

FIG. 1 Substrates for circular permutation, phasing analysis, and cyclization kinetics. DNAs with internal loops are denoted wl and those without as all-ds; 'loop DNA' refers specifically to DNA comprising the loops, denoted by open ovals. Asterisks indicate positions of ^{32}P label. a, Internal loops formed by hybridization of 182-bp *Bss*HI fragments. WI-R and WI-Y indicate heteroduplexes with purine and pyrimidine loops. Plasmids differing in the loop region contribute strands 1 and 2. b, Phasing of internal loops against sequence-directed A-tract bends. Black wedges represent phasing adaptors. Base pairs removed in a second set of molecules used for cyclization are bracketed. Angle brackets indicate the 3' ends of labelled 19-nucleotide oligonucleotides, complementary and mismatched to the bottom strand, which were hybridized with oligonucleotides extending beyond the *Nco*I site and ligated in 10-fold excess to the *Xba*I site of *Nco*–*Xba*I fragments bearing A-tracts flanked by phasing adaptors (from previously described constructs¹⁰, supplemented with two similar molecules). *Nco*I restriction provided the final products, which were gel-purified. Duplex sequences are written 5' to 3' on the upper strand.

METHODS. For circular permutation substrates, oligonucleotides were cloned into *Cla*I–*Hind*III restricted pBluescript II KS(+) (Stratagene) by standard methods. Gel-purified [γ - ^{32}P]ATP-labelled *Bss*HI fragments were mixed with 40-fold excesses of cold fragment from the other plasmid (total 50 nM in 16 mM HEPES, pH 8.0, 40 mM NaCl), boiled for 5 min and allowed to cool overnight. All-ds molecules were similarly boiled and rehybridized. The internal loop in an otherwise intact molecule was detected by *Sty*I resistance (for heteroduplexes) and *Hind*III sensitivity. For construction of phasing substrates, ligation and restriction were done according to the manufacturer's protocol (New England Biolabs).

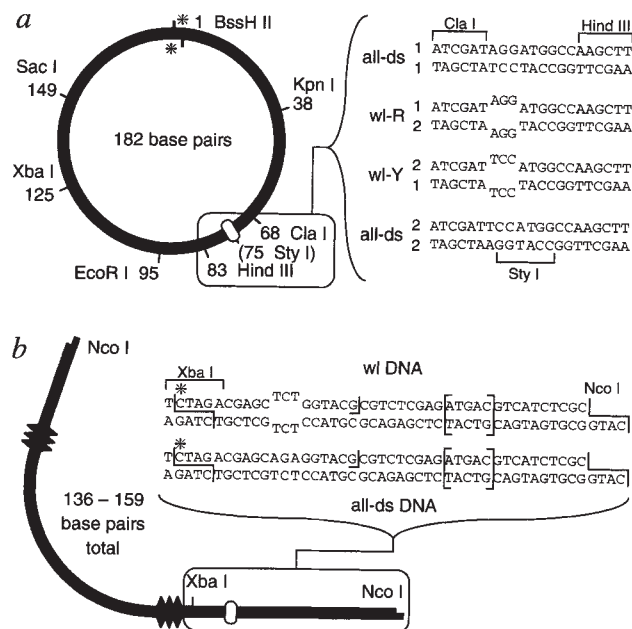


FIG. 2 Circular permutation of the flexible loop. a, Migration of wl DNA and all-ds DNA in an 8% gel. Each set of four lanes are from restriction, with the indicated enzyme, of ligated *Bss*HI fragments. Labels 1*1, 1*2, 12* and 22* indicate the strands present, with ^{32}P indicated by an asterisk, so outer lanes are all-ds and inner lanes are wl DNA. wl-Y DNA migrates behind wl-R, as shown by restriction of 182-bp wl *Sac*I and *Bss*HI bands with *Bss*HI and *Hind*III respectively (not shown). For example, *Bss*HI restriction of the top 1*2 *Sac*I band gives a 33-bp product, which must be bottom-strand labelled, and therefore must be from wl-Y. Slower migration of wl-Y DNAs is confirmed by the longer loop-bearing fragments, which result from head-to-head ligation between labelled wl DNA and excess cold all-ds DNA and are therefore uniquely labelled. All-ds head-to-head products provided length standards for calculating R_L values (apparent length divided by actual length, normalized to the extrapolated value for the loop at the end). b, Mobility of each wl DNA relative to the corresponding all-ds DNA, as a function of loop position, for 6, 8 and 12% gels. Cosine curves are fitted excluding *Cla*I mobilities. R_L values estimated for loops at the centre of wl-Y DNA at the three gel percentages are 1.09, 1.11 and 1.12, and for wl-R are 1.07, 1.07 and 1.06. This weak dependence of R_L on gel percentage may be a useful diagnostic for flexibility effects.

METHODS. Rehybridized *Bss*HI fragments (20 nM) were ligated by standard methods, giving mainly circular products, and split into six restriction reactions. Products were analysed on 30 cm $\times$ 40 cm $\times$ 0.5 mm gels (40:1 acrylamide:bis; 0.6 $\times$ TBE), run at room temperature for several hours. Size calibration (in base pairs) is indicated on the sides.

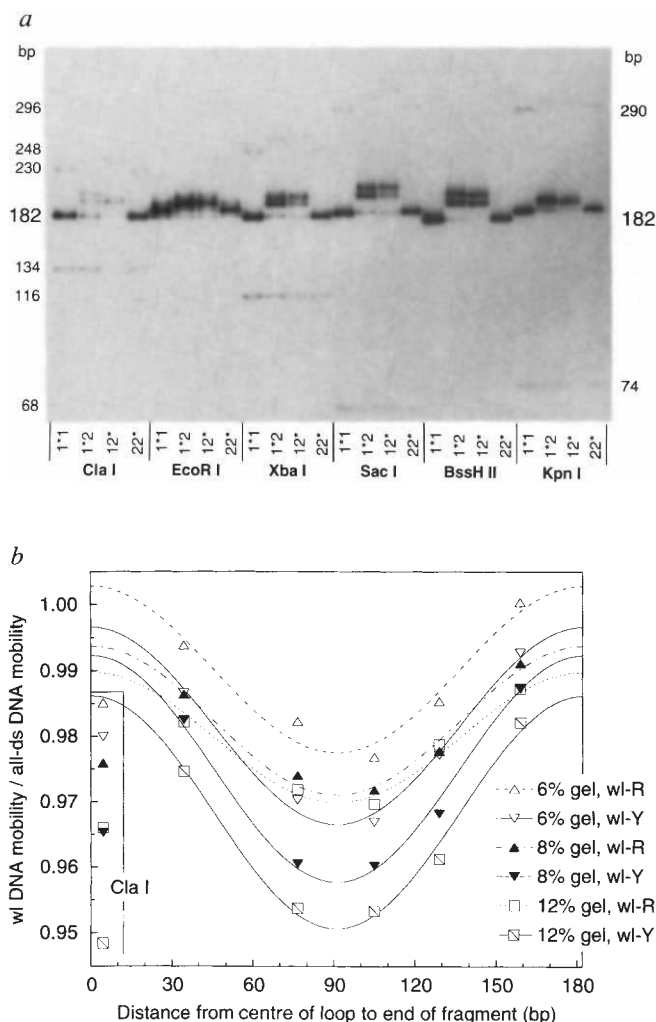
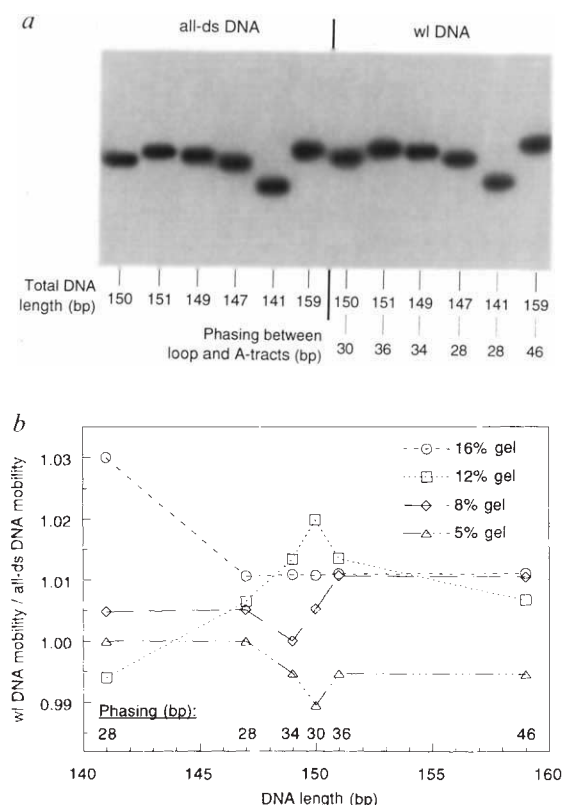


FIG. 3 Phasing of flexible loop against A-tract bends. *a*, all-ds DNAs are on the left and wl DNAs on the right of the autoradiogram of a 12% gel, with the indicated number of base pairs separating the loop from the first A of the A-tracts. The total DNA length varies independently of the phasing because of the phasing adaptors on the other side of the A-tracts. *b*, Relative mobility $\mu(\text{wl})/\mu(\text{all-ds})$ for wl versus all-ds DNA is shown as a function of DNA length and bend phasing for four gel percentages. The average $\mu(\text{wl})/\mu(\text{all-ds})$ values are 0.996, 1.006, 1.009 and 1.014 for 5, 8, 12 and 16% gels, whereas the average R_L values for all the DNAs, relative to unbent DNA, are 1.33, 1.84, 2.43 and 2.53 for the four gel percentages.

METHODS. Gel electrophoresis was performed as in Fig. 2, except 18 cm $\times$ 16 cm $\times$ 1.5 mm gels were used, thermostated at 20 °C in a Hoefer SE600 apparatus. R_L values are relative to an *MspI* digest of pBR322.



has little effect on the electrophoretic mobility of curved DNA molecules, raising questions about the assumption that curvature and flexibility behave equivalently in that they affect mobility only through average DNA conformation^{21,22}. Curved and flexible molecules differ in that the latter provide no restoring potential to adopt a bent shape when straightened by passage through a tube in the gel.

In contrast to its subtle effects on electrophoresis, loop-induced flexibility dramatically affects the cyclization probability (J factor) for wl compared with all-ds DNAs (Fig. 4). J , the effective concentration of one properly aligned and helically phased DNA end around the other end, is strongly dependent on the number of DNA helical turns²³. The internal loop removes this dependence, demonstrating that loop DNA is much more torsionally flexible than dsDNA. In Fig. 4*a*, cyclization of 141-bp all-ds DNA (a half-integral number of helical turns) is very slow, whereas 141-bp wl DNA cyclizes 100 times more readily. Figure 4*c* illustrates the cyclization probability for the set of molecules and suggests that the optimum length for cyclizing wl DNA may be slightly longer than for all-ds, corresponding to about 1 bp of loop-induced unwinding. Furthermore, if the loop were simply a torsional swivel, we would have $\langle J \rangle_{\text{wl}} = \langle J \rangle_{\text{all-ds}}$, so the observed overall cyclization enhancement in wl DNA indicates increased bending flexibility in the loop. For example, although all-ds 147-bp DNA has a near-integral number of helical turns, wl DNA cyclization is still enhanced threefold (Fig. 4*b*). Torsional flexibility randomizes bend direction, so there are no obvious effects of loop position relative to the A-tracts.

Preliminary analysis of the cyclization kinetics gives estimates for C_{loop} and P_{loop} , the loop DNA torsional modulus and persistence length; the calculations ignore writhe, which is reasonable for small circles, and assume harmonic bending and twisting energies for loop DNA. The length dependencies (Fig. 4*c*) of $J_{\text{all-ds}}$ and J_{wl} give $C_{\text{ds}} = 2.1 \times 10^{-19}$ erg cm, within the range of previous measurements⁹, and $C_{\text{loop}} \approx 3.2 \times 10^{-21}$ erg cm. This corresponds to a twist standard deviation of 66° for the loop as

a whole, compared with 8.1° for 3 bp of dsDNA. The 66° estimate is independent of the number of base pairs in the loop, but if 5 bp are actually melted we obtain $C_{\text{loop}} \approx 5.3 \times 10^{-21}$ erg cm, in agreement with an estimate from denaturation kinetics²⁴. We estimate loop bending flexibility (P_{loop}) assuming that cyclization enhancement reflects the bending free energy difference for cyclization with and without the loop, and that bending in the circle is distributed so as to minimize $\Delta\Delta G_{\text{bend}}$ (Fig. 4*c*). This yields $P_{\text{loop}} = 10.5$ bp (36 Å), corresponding to an r.m.s. loop bend angle of $43 \pm 5^\circ$ (error estimate corresponds to an uncertainty of $\pm(\log 1.5)$ in $\langle \log(J_{\text{wl}}/J_{\text{ds}}) \rangle$), versus 11° for 3 bp of dsDNA; again, the predicted loop bend angle is essentially independent of the assumed loop size. These rough estimates will be improved using Monte Carlo simulation^{9,20,25,26}.

In conclusion, our experiments show that DNA cyclization kinetics is the preferred method for studying localized flexibility. Cyclization allows detection and measurement of independent changes in torsional and bending flexibility, whereas electrophoretic calibration curves^{13,19} underestimate the average bend due to a flexible loop. Bends in the phasing assay have a complicated geometry and are usually apparently larger than in circular permutation, causing disagreements between quantitative results from the two experiments; given the insensitivity of electrophoresis to flexibility, such disagreements should not be used to infer flexibility. This does not, however, imply that proteins cannot induce flexibility—in fact, isotropic flexibility may often go undetected because of its modest influence on mobility.

Finally, we can deduce from our results that bending can help induce base-pair separation, as suggested theoretically¹². We have shown that unpaired DNA is easier to bend or twist than duplex DNA, and this means that unpairing must be favoured by bending or torsional stress; the free energy cost of unpairing is compensated for by the consequent decreased bending or twisting free energy. The relationship parallels thermodynamic cycles requiring improved binding of bending proteins to prevent DNA¹⁰ or of unwinding ligands to negative supercoils. This supports another biological role for DNA looping (in the sense

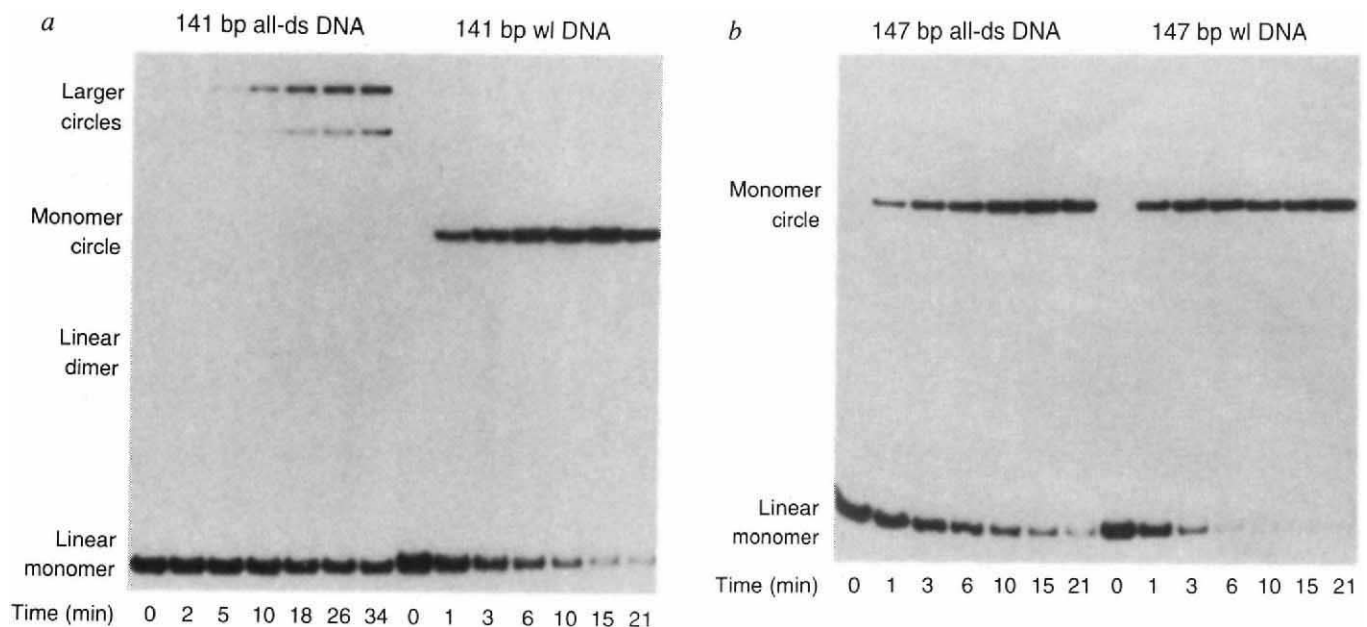


FIG. 4 Effects of an internal loop on DNA cyclization kinetics of phasing constructs. *a*, Autoradiogram showing ligation of 141-bp all-ds and wl DNA. Bimolecular and monomolecular products are indicated. *b*, Cyclization of 147-bp all-ds and wl DNA. *c*, J for all-ds and wl DNA versus DNA length. Data are fitted to $J = J_{\max} \sum \exp[(\tau - \tau_0)^2 / 2\sigma_{\text{app}}^2]$, where $J_{\max}$ is the optimum cyclization probability, τ is the total twist (bp), σ_{app} is the apparent standard deviation of total twist, and the summation runs over τ_0 corresponding to integral helical repeats. σ_{app} in radians was derived from the helical repeat, 10.58 bp per turn, estimated from the DNA length for maximum $J_{\text{all-ds}}$ and the A-tract helical repeat (10.33)¹⁸. The torsional modulus for dsDNA is $C_{\text{ds}} = kT/\sigma_{\text{ds}}^2$, where the base-pair rise l is 3.4 Å, $\sigma_{\text{ds}}^2 = \sigma_{\text{app}}^2(\text{all-ds})/N$, and $N \approx 147$ is the total number of base pairs²³. Similarly, $C_{\text{loop}} = kT/\sigma_{\text{loop}}^2$, where $\sigma_{\text{loop}}^2 = \frac{1}{3}[\sigma_{\text{app}}^2(\text{wl}) - (N-3)\sigma_{\text{ds}}^2]$. Loop persistence length is calculated from the change in average cyclization free energy, $\Delta\Delta G = 2.3RT(\langle \log_{10} J \rangle_{\text{wl}} - \langle \log_{10} J \rangle_{\text{all-ds}}) = 1.35 \times (1.42 - 0.46) \text{ kcal mol}^{-1} = 1.3 \text{ kcal mol}^{-1}$, assumed equal to the change in bending free energy, $\Delta\Delta G_{\text{bend}}$. DNA bending free energy is $RTP(\Delta\theta)^2/2L = 0.014 (\Delta\theta)^2/2L$, where $\Delta\theta$ is the bend angle over L bp and the persistence length $P = 150$ bp for dsDNA³⁰. Here $\Delta\theta = 252^\circ$ because six A-tracts contribute 108° , giving

$$\Delta\Delta G_{\text{bend}} = \left[\frac{0.014}{147 \text{ bp}} (252^\circ)^2 \right] - \left[\frac{0.014}{144 \text{ bp}} (252^\circ - x^\circ)^2 + \frac{RTP_{\text{loop}}}{2 \times 3 \text{ bp}} (x^\circ)^2 \right] = 1.3 \text{ kcal mol}^{-1}$$

where x° is the bending in the loop. Setting $\partial\Delta\Delta G_{\text{bend}}/\partial x = 0$ gives x in

terms of P_{loop} , which we use in the above equation to solve iteratively for P_{loop} . The root-mean-square bend angle in linear DNA is $\sqrt{2 \times 3 \text{ bp} \div P_{\text{loop}}}$ radians²⁰.

METHODS. Cyclization reactions were carried out and analysed as reported¹⁰, except ligation buffer contained 10 mM MgCl₂. T4 ligase concentrations were 333 or 1,000 units ml⁻¹. Values for $\log_{10} (J_{\text{wl}}/J_{\text{ds}})$ are reproducible to ± 0.5 .

of bending dsDNA around to approach itself): in addition to bringing protein ligands together, looping and the resulting bend may help unpair DNA. Bending-induced unpairing may be a driving force for open complex formation in transcription²⁷ and for unwinding the origin of DNA replication^{28,29}. □

Received 12 October; accepted 7 December 1993.

1. Crothers, D. M. & Steitz, T. A. in *Transcriptional Regulation* (eds McKnight, S. L. & Yamamoto, K. R.) 501–534 (Cold Spring Harbor Laboratory Press, New York, 1992).
2. Drew, H. R. & Travers, A. A. *J. molec. Biol.* **186**, 773–790 (1985).
3. Goodman, S. D. & Nash, H. A. *Nature* **341**, 251–254 (1989).
4. Wu, H.-M. & Crothers, D. M. *Nature* **308**, 509–513 (1984).
5. Zinkel, S. S. & Crothers, D. M. *Nature* **328**, 178–181 (1987).
6. Salvo, J. J. & Grindley, N. D. F. *Nucleic Acids Res.* **15**, 9771–9779 (1987).
7. Hodges-Garcia, Y., Hagerman, P. J. & Pettijohn, D. E. *J. biol. Chem.* **264**, 14621–14623 (1989).
8. Kotlarz, D., Fritsch, A. & Buc, H. *EMBO J.* **5**, 799–803 (1986).
9. Crothers, D. M., Drak, J., Kahn, J. D. & Levene, S. D. *Meth. Enzym.* **212**, 1–29 (1992).
10. Kahn, J. D. & Crothers, D. M. *Proc. natn. Acad. Sci. U.S.A.* **89**, 6343–6347 (1992).
11. Kerppola, T. K. & Curran, T. *Science* **254**, 1210–1214 (1991).

12. Ramstein, J. & Lavery, R. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7231–7235 (1988).
13. Thompson, J. F. & Landy, A. *Nucleic Acids Res.* **16**, 9687–9705 (1988).
14. Levene, S. D. & Zimm, B. H. *Science* **245**, 396–399 (1989).
15. Zimm, B. H. & Lumpkin, O. *Macromolecules* **26**, 226–234 (1993).
16. Lerman, L. S. & Silverstein, K. *Meth. Enzym.* **155**, 482–501 (1987).
17. Werel, W., Schickor, P. & Heumann, H. *EMBO J.* **10**, 2589–2594 (1991).
18. Drak, J. & Crothers, D. M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3074–3078 (1991).
19. Zinkel, S. S. & Crothers, D. M. *Biopolymers* **29**, 29–38 (1990).
20. Koo, H.-S., Drak, J., Rice, J. A. & Crothers, D. M. *Biochemistry* **29**, 4227–4234 (1990).
21. Olson, W. K., Marky, N. L., Jernigan, R. L. & Zhurkin, V. B. *J. molec. Biol.* **232**, 530–551 (1993).
22. Zhurkin, V. B., Ulyanov, N. B., Gorin, A. A. & Jernigan, R. L. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7046–7050 (1991).
23. Shore, D. & Baldwin, R. L. *J. molec. Biol.* **170**, 957–981 (1983).
24. Crothers, D. M. & Spatz, H. C. *Biopolymers* **10**, 1949–1972 (1971).
25. Hagerman, P. J. & Ramadevi, V. A. *J. molec. Biol.* **212**, 351–362 (1990).
26. Levene, S. D. & Crothers, D. M. *J. molec. Biol.* **189**, 61–72 (1986).
27. Travers, A. A. *CRC Crit. Rev. Biochem.* **22**, 181–219 (1987).
28. Gille, H., Egan, J. B., Roth, A. & Messer, W. *Nucleic Acids Res.* **19**, 4167–4172 (1991).
29. Hwang, D. S. & Kornberg, A. *J. biol. Chem.* **267**, 23083–23086 (1992).
30. Liu-Johnson, H.-N., Gartenberg, M. & Crothers, D. M. *Cell* **47**, 995–1005 (1986).

ACKNOWLEDGEMENTS. This work was supported by an NIH grant (D.M.C.) and an NIH postdoctoral fellowship (J.D.K.).

Lift-off for bovine genome project

A linkage map for of the cattle genome points to mechanisms of the reassortment of mammalian genes while a new technique raises questions about the structure of gene-starts in chromatin.

THE genome of the cow (otherwise called the bovine genome) is this month's claim on public attention in *Nature Genetics* for several reasons. For one thing, bovids are also mammals. For another, they are animals of great economic importance in which advantageous and distinguishable traits command high prices at rural auction sales, with the result that the fate of individuals is closely watched by fond owners. Then the practice of artificial insemination and the relatively short time-lag in cattle between birth and sexual maturity means that hugely informative pedigrees can readily be available if the record-keeping is reasonably good.

So it is only natural that a detailed genetic map of the bovine genome should be a by-product of state-sponsored cattle-breeding (a similar map from the US Department of Agriculture has been published in the February issue of *Genetics*). Almost predictably, Australia, and the Commonwealth Scientific and Industrial Research Organization (CSIRO), seems to have taken the lead in this project. Six out of 23 authors of the map (*Nature Genet.* 6, 227–237; 1994) are from the CSIRO laboratory of the Tropical Animal Production Division at the University of Queensland at Brisbane. But collaborating laboratories span the world, from Texas A&M, the International Laboratory for Research on Animal Diseases (ILRAD) at Nairobi and the University of Wisconsin at Madison (all predictable) to Trinity College, Dublin, the University of Georgia, the Hebrew University of Jerusalem and Eragnische Technische Hochschule (ETH) at Zurich (in varying degrees, less so).

The product of the collaboration is, at this stage, a linkage map only. In other words, the collaboration has considered a subset of 202 conveniently polymorphic loci of the bovine genome out of the more than 300 genetic so far assigned to chromosomes and has sought to put them in topographical order by analysis of the frequency with which putative nearest-neighbour variants are linked in different individuals. The result is not so much a map, as the starting-point for the construction of a map. It is both sobering and encouraging that the mapping of the human genome was only a little further advanced at the beginning of this decade — sobering, because the collaboration illustrates again the difficulty of making a map of any kind when the topography is

devoid of landmarks, and encouraging because, in the human case, there has since been a ten-fold refinement.

Cattle-breeders will naturally await the more detailed maps that lie ahead, but others will be surprised at the guesses at the mechanism of mammalian speciation that the authors have been able to make by comparing their crude linkage map for cattle with the much more detailed maps now available for mice and people. The issue is simply put. At least to a first approximation, there is an uncanny resemblance between the genes of different mammals. Human genes are usually homologous with those of the mouse (whence the case, sometimes put, for making the Murine Genome Project a mode for the human version). So is the rudimentary bovine map more like that of mouse or Man?

Cattle have had 60 chromosomes to the genome (29 pairs of autosomal chromosomes and an X and a Y), compared with 46 in people. Given the apparently explosive radiation of very different super-families of mammals in the Cenezoic, it would be folly to expect the any one of the bovine, human and murine genomes to be ancestral to the other two. At least on the resolution of the new bovine map, the three species are equally alike (or different). In each of the three two-way comparisons now possible, while some stretches of chromosomes appear to have been simply translocated, other groups of linked genes seem to have been shuffled without rhyme or reason. But that in itself is a challenge; it is tempting to believe that there must be a pattern of some kind.

Meanwhile, there is also a potentially important improvement of sequencing technique to report, based on an observation of the non-random distribution of CG dinucleotide pairs in the mammalian genome. If it were that nucleotide sequences were strictly random, the chance that any pair of consecutive nucleotides would be a CG would be 1 in 16, but in reality the occurrence of CG pairs is far from random. Mostly they crop up once every 50 to 100 base-pairs and are heavily methylated, but they occur much more frequently (1 in 10 base-pairs or so) upstream of the 5'- ends of mammalian genes, when they are usually not methylated. So much has been demonstrated during the past decade by Adrian Bird and successive colleagues at

the University of Edinburgh. Now *Nature Genet.* 6, 236; 1994), Bird's group shows how the non- random distribution of CG pairs may be used in at least two ways in large-scale sequencing efforts.

The starting-point is a technique for separating short lengths of DNA containing methylated and unmethylated CG pairs, based on the use of a polypeptide (MBD) which is itself a portion of a protein associated with and isolated from rat chromosomes and which has the property of binding to symmetrically (or equally) methylated CG pairs. The goal is to isolate the CG "islands" which Bird's group (and others) have shown to span the upstream and early-exon parts of 60 per cent or so of mammalian genes. So why not incorporate the MBD peptide into a retention column, chop up the genomic DNA into fragments with a restriction enzyme cutting at sites that rarely occur within CG islands and use the retention column to fragments with methylated CG pairs?

The goal is to isolate the estimated 45,000 CG islands in the human genome, each typically 1,000 base-pairs long. The retention column works both as a means of removing non-island (or methylated) DNA as a means of fractionating the island sequences when they are artificially methylated after being segregated as a mixture. One potential prize is that the relatively simple task of sequencing 45,000 CG-islands, each roughly a kilobase in length, will yield as many distinctive genetic markers. But because the islands span both the promoters of the genes to which they correspond and at least the beginning of the transcribed mRNA, they should be distinctive means of isolating intact cDNA molecules, not the partial sequences usually produced. E

Obviously, there are snags. The restriction enzymes chosen so as to cut DNA only outside CG islands are not as specific as they might be, presumably because their target cutting-sequences occasionally occur within islands (which helps to explain why the average length of the fragments is less than expected. And some of the fragments are from ribosomal or mitochondrial DNA. It is also, of course, the plain truth that, if unmethylated CG-islands are associated with active genes, 40 per cent of the genes in the human genome will not be so identified. But that, it may be thought, is a lot better than nothing. □

Science parks as a force in employment

Science parks are all the rage, not just as locations for new industries but as means of making universities grander places. What follows is a symposium of sometimes fulsome opinions, often by managers of science parks, of their advantages to researchers.

Bordeaux Technopolis and sustainable development

Yann Couvidat

THE very concept of sustainable development means thinking in terms of adaptation. To be sustainable means also to be adaptable to the constant evolution of the environment. Thus, to think in terms of sustainable development, one has to find good answers not only to the problems of the moment but also to the problems of tomorrow.

In September 1994 the Bordeaux Technopolis project welcomes participants from the world of science parks to a conference on the theme of 'sustainable development and science parks'. This has been an underlying theme of developments at Bordeaux.

Durability and sustainable development do not mean no economic activity and no industry, and being worried about durability does not mean forgetting the development necessary for the evolution of society. Development implies at least two phases. First, emulation, which is still too often a simple copy of the industrialized countries' development. Then, autonomization, with not only the improvement of products from the industrialized countries and the means of producing them, but also new products, so that developing countries can compete with developed countries. This is where science parks have an important role to play. At the same time, the very local development of a science park must not mean irresponsibility and distortion in the local economies and urban structures.

We all know that the twenty-first century will be an urban century. Then, we have to think about development and sustainable development through the city, and to think about the city through policies for evolution. This is one of the lessons the Bordeaux Technopolis can help to illuminate. It may well be one of the oldest science park projects in the world. Its origins go back to 1963, but the site became operational only in the early 1990s.

So, what did happen in 1963?

French local and national authorities decided to build a campus intended to be the largest in Europe; it could not itself develop without a companion area where enterprises could develop and grow. This kind of planning, which has been going on

for more than 30 years, with its share of local accidents and political incidents, has produced not only the largest and one of the greatest campuses in Europe, but a fairly efficient urban network of science parks; the webs of this network are now building the city of tomorrow.

The Bordeaux Technopolis is certainly not the most developed, and it is not even the best planned. But it is, undoubtedly, the one that has received the most attention from the largest number of people over the longest period of time. This approach can sometimes have its drawbacks. For consensus is not always reached on many issues, and debates are and have been the constant backdrop to the policies on science parks in that city.

Out of that long gestation came a project that is now mature and very specific. The project is linked to the best of the local universities, built on local strengths, managed with understanding of local conditions, and yet involving the whole area,



without overlooking one side of the city, or one area of excellence, or setting aside any one path of development. After those hesitant years, the best result may be the scope of the thinking that has been lavished on the project.

What are the local strengths? Bordeaux's immediate fame rests on wine. But the trouble is that the vineyard somewhat overshadows the other qualities of the area.

Bordeaux Technopolis is in fact a city of technical advance and science. First, Aerospatiale, SEP (European Company of Propulsion), Dassault and so on: 20,000 people are employed in an area with no equivalent in Europe in aeronautics,

space and ballistics. Ariane rocket engines are built here. Bordeaux Technowest, close to the international airport, is networking these activities, making sure that they will develop by taking into account progress in science and technology, and today's needs as well as tomorrow's for space and urban planning.

Bordeaux ranks third among medical and health science areas in France. From classical drug manufacture to the most advanced molecular implant engineering, more than 6,000 employees work in health industry companies. It is thus natural to have devoted part of the Bordeaux Montesquieu site to life sciences and techniques. However, as we shall see, it is not the only peculiarity of Bordeaux Montesquieu.

Electronics is also very strong. At first it went hand in hand with the aeronautics industry, but it quickly developed as an area of growth in its own right. IBM, Motorola, Siemens — the big names are here. But over and above these, the strength of Bordeaux lies in small and medium enterprises that have embraced quality production. COM1 and Lectra Systems, for instance, are two of the principal emerging ones. But as well as research and development, there is also high quality production — as Pioneer and many others witness.

Bordeaux Hauts-de-Garonne, on the upper right bank, will be the home of what the French call middletech and logistics. It is so far the less developed aspect of the technopolis.

The technopolis wove its network into those existing strengths, to make them more interconnected and to evolve into new directions — in short, cross-synergy. From the technologies of composites needed to serve aeronautics, to health sciences, the Institute of Advanced Materials, located on Bordeaux Montesquieu, tests every day the requirements of cross-synergy.

Bordeaux Unitec is located on the university campus and weaves through the neighbouring cities to unite local strengths, create synergies and make them prime forces for local development. Promising fields will be chemistry, mathematics and their intermixture with other sciences.

Bordeaux Technopolis has developed five sites in all. And, by respecting local strengths, forecasting or trying to forecast urban development, and building on the

PARK OWNERSHIP

Type of ownership	North American parks		Other parks		Worldwide parks	
	Parks	%	Parks	%	Parks	%
University	37	34	13	15	50	26
Government	15	14	29	33	44	22
For profit	19	18	9	10	28	14
Private not-for-profit	22	20	4	4	26	13
Joint venture	15	14	34	38	49	25
Totals	108	100	89	100	197	100

The type of park ownership inside and outside North America is quite different. In North America, universities own a large proportion of parks, with other types of ownership evenly divided. Parks outside North America are largely owned by joint ventures and government entities. Almost half of parks worldwide are owned by universities and government entities.

WORLDWIDE PARKS ESTABLISHED BY YEAR

1950-59	5	1985	34
1960-69	11	1986	39
1970-74	12	1987	31
1975-79	9	1988	24
1980	9	1989	25
1981	4	1990	21
1982	10	1991	9
1983	25	1992	8
1984	36	1993	3
Planned		29	

Source of data as for table on page 172.

specific local ecology, it is slowly succeeding in showing how sustainable development can be achieved through development of local enterprises. But it would have been impossible to develop such a scheme without taking into account the necessary growth of the city and respect for the environment. On the edge of the largest forest in Europe, one could not develop a science park without considering water management, trees and links to existing infrastructure that does not destroy local ecology.

"Bordeaux Montesquieu is emerging as one of the first ecological technology and science parks around the world. "In saying this, the Bordeaux Technopolis team does not only mean having companies dealing with ecological issues, or laboratories checking the ecological balance of the site. Located in a forested area, and on the edge of the vineyards, it had to be ecologically aware during the planning process. Ten years ago, when the actual planning took place, this was not a common preoccupation of science park planners. Bordeaux Montesquieu is the only science park in the world that deals with surface water by using its roads as waterways and filters. Surface water is collected by the roads, filtered by the road surfaces, then lagooned by the lake. The same kind of awareness has been shown when building, respecting most of the local trees and planting new ones when necessary.

The lessons from Bordeaux may be applicable elsewhere. Science parks and technopolitan projects nowadays need to be woven into the texture of urban complexity, taking into account its dynamics and evolution and trying to promote the emergence of new technologies, without impeding tomorrow's development. The

cities have traditionally been the places where wealth accumulation takes place, where trade and exchange take place. The city is also the place where decisions are taken. It is the setting for everyday life. It is also where most science parks have been developed around the world.

Science parks are tools for sustainable development, for they help to bridge the gap between rich and poor nations. But they can do more: through cautious and necessary planning, they can help cities to grow, without forgetting the requirements of tomorrow.

Yann C. Couvidat is Executive Officer of the International Association of Science Parks, at Centre Montesquieu, Bordeaux Technopolis, 33650 Martillac Cedex, France.

Research Triangle Park

Jeanne P. Brewer

JANUARY 9, 1994 was a milestone in the history of Research Triangle Park in the state of North Carolina. On that day, the park marked its thirty-fifth anniversary. That, in retrospect, is when North Carolina embraced new technologies, new products, new strategies and new thinking all at once.

After 35 years of growth, the 6,800-acre Research Triangle Park touches the lives of people throughout the world. Whether it is from viewing a football game played on Astroturf, an artificial turf product invented in the Park; using products that carry a UPC code and were scanned at a local store, a process invented in the park; or using one of the many medical products created by the companies there, products

such as AZT, Exosurf, Adenoscan, Zovirax or Zantac, the park's effect is truly global. Even employment in the park has led the growing trend to globalization. Since the first internationally based company, Burroughs Wellcome Co., located itself in the park in 1972, employment has been evenly balanced between nationally and internationally based companies. Companies from Britain, France, Germany, Japan and Sweden grow side by side with companies that moved from New Jersey, California and Connecticut.

This momentum creates the jobs and challenges that drive companies to succeed and grow. The park's 65 research companies and 55 service companies employ 34,000 people. People working in the park live throughout the Triangle area and hail from a wide range of scientific disciplines, cultures and countries. Not only has this growth brought jobs within the park, but its influence has spread wealth to the whole area. A 1989 study showed that an additional 20,000 jobs have been created that would not have existed without the park. That net effect — 54,000 jobs by 1989 figures and some 60,000 by 1993 estimates — represents more than 10 per cent of the entire workforce of the Raleigh-Durham region.

Research Triangle Park has met the expectations of its founders in creating



employment for the state and the region, and for graduates of the Triangle universities. We have been able to reverse the brain drain, the trend that 35 years ago saw many of the brightest graduates leaving North Carolina to find suitable employment. Today, the Research Triangle Park is a net importer of PhDs and influences life in the Triangle in many ways. The Raleigh-Durham International Airport recently received confirmation of a route to London, based largely on the number of international companies in the park and across the state. An array of conventions choose the Raleigh-Durham area as a venue because of the presence of the park, and groups from around the world come to see a research park that looks like a true 'park', with green spaces and acres of trees.

With each passing year, the park receives increased attention. Magazines such as *US News & World Report*, *Location*, *Site Selection* and *Fortune* have kept a consistent spotlight on the Triangle area. *US News & World Report* has highlighted the Triangle universities regularly in its rankings of universities. In 1992, the first year in which *Fortune* magazine included the Triangle area in its rankings, Raleigh-Durham was ranked the sixth best area in the United States for global competitiveness. The momentum picked up as *Location* ranked the Triangle in the top six markets in which to locate a business, and *US News* ranked the area as among the top 20 markets in which to buy a home. *Business Week* then covered North Carolina and the growth of the I-85 corridor, including the Research Triangle Park, noting it as an area to watch for continued growth in high technology and business. *Money* magazine ranked the area in the top five best places to live, *World Trade* magazine rated the Triangle in the top five locations for international companies, and *Places Rated Almanac* said the region was the sixth best place to live in the United States. Best of all, however, was the 15 November 1993 ranking by *Fortune* that listed the Best cities for business: Where to find America's knowledge workers. "Raleigh-Durham, no.1 on our list, is a smart choice" was the description on its cover.

For those who live and work in Triangle, the ranking came as no surprise. Not only are there the larger companies in the park and around the Triangle, but also the small, start-up companies and mid-sized, growing companies. Since 1959 more than 358 technology-orientated start-up companies have been created and survived. And the park's newest tenant, Medco Research, a small pharmaceutical firm, has just announced that it has received Food and Drug Administration approval for its new drug Adenoscan.

The Research Triangle Park today is a no less exciting place to be than it was 35 years ago. The tradition of commitment, cooperation and excellence continues as the Research Triangle Foundation of North Carolina — the owners and developers of the park — strives to bring the park to the attention of new companies worldwide. The growth of the park continues at a steady pace, with the remaining 2,500 acres of rolling, wooded sites gradually being sold. In short, the park continues to meet the goals of its founders: to bring new, technology-based jobs to North Carolina and provide a base of employment for its citizens and graduates.

Jeanne P. Brewer is Director of Communications, Research Triangle Foundation, 2 Hanes Drive, PO Box 12255, Research Triangle Park, North Carolina 27709, USA.

Cambridge Science Park

Lindy Beveridge

THE Cambridge Science Park celebrates its twenty-fifth anniversary next year and now boasts 73 tenant companies, most of them locally born, small and newish, and operating mainly in niche markets worldwide. They are intermixed with UK subsidiaries of multinational companies based in the United States, Japan, France, Germany, Sweden, Finland and the Netherlands. All attract high-quality applicants when they advertise jobs, and as the competition for good scientific and technical staff is fierce, the scope for selection from a good list of applicants is an important reason why they want to be at the park.

The decision to develop the Cambridge Science Park was taken by Trinity College, Cambridge, its owner, in 1970 in response to a government initiative to encourage stronger links between universities and industry in the United Kingdom, and to the Mott Report which recommended the creation of a special location in Cambridge to encourage the growth of high-technology industry. The creation of 2,000 new, highly skilled jobs was one of the targets proposed for this development. Twenty-four years later, more than 3,500 new jobs have been created and the park's success has been influential in encouraging the development of other science parks in the United Kingdom and elsewhere.

Most companies at the park concentrate their financial resources on recruiting and keeping good technical staff and equipping them with the best laboratory facilities. Services needed for manufacture and marketing of products are generally bought in from local sources and most concentrate their activity on research and development, testing and marketing rather than on manufacture.

Conditions of work are excellent. Park

companies occupy purpose-built accommodation in green and pleasant surroundings, and are surrounded by like-minded people, as the planning use clauses for the park restrict its tenants to only three kinds of activity — applied research, development of products and services which require a continuous input of research and development, and specialist services.

Social and conference facilities are especially important to smaller companies which may have only a small staff and limited work space. There is a pleasant social facility — the Trinity Centre — where park tenants can eat and meet throughout the day, with squash courts nearby and a variety of social clubs and other activities such as keep-fit classes and blood donor sessions. In general, park companies do not poach each other's staff,



but there is some scope for changing jobs within the park community and this is the place to hear about new opportunities.

In the past, and especially in the early days of the park's development, University of Cambridge scientists often regarded scientific staff in the park companies as being involved in rather unchallenging work. But there has been a sea change in attitudes over the years, largely thanks to a special research fund available to park



The 'Phase 3 Lake' and saplings in Cambridge's park.

companies for collaborative research. This has led to many fruitful working relationships between park scientists and university staff. A nexus of contacts now exists, with university scientists acting as consultants, advisory board members or directors of park companies as well as having more informal friendships with them. In the past ten years, an increasing number of companies founded by university scientists has moved to the park, bringing valuable opportunities for contact with them and facilitating the sharing and overlap of the university culture with that of the local high-tech industry for which the park acts as a flagship.

Park companies rely on the excellence of their science for their competitive position in world markets and they are often among the top five in their fields. They need the best microbiologists, chemists, electronics specialists and physicists and, often, people who can combine traditional disciplines in new hybrids for such developments as biosensors. In many cases, they seek to recruit young, talented scientists who are up to date with recent developments in their field and have had a year or two of industrial experience. Communications skills are important, not only for commercial purposes but also because companies at the park have unusual opportunities to meet visitors from all over the world and are often asked to explain or demonstrate their work to government ministers and civil servants, politicians, heads of universities and research institutes as well as royalty.

From the start, Trinity College set out to protect the park companies from unnecessary bureaucracy. It has an agreement with the local planning authorities that enables prospective tenants to know from their first meeting if they will be acceptable in terms of the park's planning restrictions, and the park is run by a loose group of professional specialists from local companies who report direct to Trinity College and to whom the park tenants have direct access to solve problems quickly as they arise. This 'no bureaucracy' system of running the park — it has no central office and only two employees (the gardeners who look after its landscape) — also ensures direct contacts with Trinity College and university departments. Cambridge's university departments and laboratories have long operated a very simple system for dealing with requests for use of their facilities by other scientists, into which the park companies are easily accommodated. Staff also have access to libraries and to seminars and lectures by simple request direct to lecturers or through Trinity College.

It is easy to see a science park as being more about high-tech equipment and products than about people. Cambridge Science Park had a scientist, Dr John Bradfield, as its first director and the tradition

has been carried on in the appointment of his successor, Dr Jeremy Fairbrother. From the start, the conviction that the only successful way to encourage collaboration between park companies and university scientists was through creating opportunities for individuals to meet and work together voluntarily, on a basis of mutual liking and respect, has informed Trinity's approach to industry/university links at the park. From time to time it holds special social events to which both staff from park companies and university scientists are invited and arranges or facili-

tates many informal contacts. Both university staff and science park companies have their own differing daily agendas. Through Trinity, a large college with the finest traditions of scientific achievement and headed by the present president of the Royal Society, both communities can hear news of the other and build the contacts they desire to their mutual benefit.

Lindy Beveridge is a consultant to the Cambridge Science Park, and is based at 27 Alpha Road, Cambridge CB4 3DQ, UK.

Taiwan's park attracts researchers home

Angela Yao

TAIWAN'S science-based industrial park in Hsinchu has been a strong magnet for drawing Taiwanese researchers and entrepreneurs back to their home country. Since its establishment in 1980, the park has lured back more than a thousand engineers and scientists, many of them from Silicon Valley in the United States. The park is now entering its third phase of development which will focus on biotechnology rather than the electronics and telecommunications industries that have made the park so famous, and the park is keen to attract people and companies in biotechnology and the biosciences from around the world.

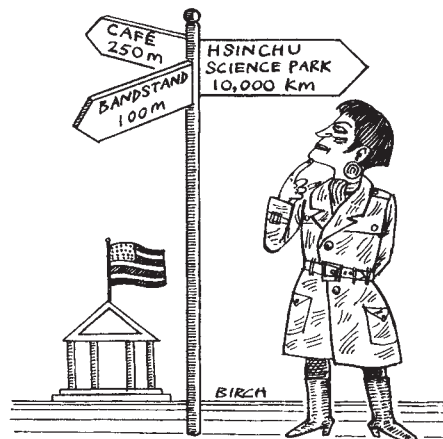
The park was established by the Taiwanese government to develop high-tech industry and provide a venue for the many talented overseas Chinese engineers and researchers. Although there are other industrial parks or zones in Taiwan, Hsinchu is the only high-tech park. Like Silicon Valley, the park is based on the principle of collaboration between scientific research and entrepreneurship, but, unlike most other such parks in the world, the Hsinchu park is planned, developed and managed by the government, or more specifically, by the National Science Council, which has a traditional role as a funding agency for academic research.

The park is located in Hsinchu, approximately 80 km south of Taipei. The site was chosen for its easy accessibility to an international airport and harbours and the plentiful technological resources nearby, including two national universities and a government-sponsored industrial technology research institute.

Since its inception, the park has received more than US\$400 million from the government, mainly for acquisition and development of land and construction. The developed land now totals 400 hectares and is divided into distinct industrial, residential and recreation areas. Another 200 hectares is being developed for the third phase expansion of the park.

To help companies to set up operations, the science park has built standard factories for rent. As an alternative, a company can also rent land to build a custom-designed factory to meet its own needs. To attract managers and engineers from overseas to the park, residential units have been built, along with recreation facilities, restaurants, and a man-made lake to provide a pleasant living environment. In addition, a unique school providing a bilingual curriculum has been set up. The school is an important attraction for the returning expatriates.

Incentives such as tax credits to encourage automation and human resource development and low-interest loans from the



government bank are being provided to companies in the park. To stimulate innovation, the park administration provides grants for innovative high-tech research and development projects to develop strategic products and components. It also selects the ten companies that put the most effort into research and development each year, and presents annual innovative product awards. Manpower training programmes sponsored by the administration in cooperation with local universities ensure a constant supply of engineers in the areas of integrated circuits, optoelectronics and integration of computers and communications. Similar programmes will be introduced in bio-science for the next phase of growth.

In spite of the threat of worldwide

recession, the park's companies continue to show an impressive record of achievement. The total number of companies operating in the park has now reached 150, of which 78 per cent are locally owned and the rest foreign-owned. Combined annual sales have grown by an impressive 48 per cent to US\$5 billion and aggregate investment by 8 per cent to US\$2.5 billion.

Six major industries are represented in the science park. In order of the number of companies engaged in each, they are integrated circuits, computers and peripherals, telecommunications, optoelectronics, precision machinery and advanced materials, and biotechnology. The number of integrated circuit manufacturers, which now total 43, has for the first time overtaken the number of computer-related companies to become the largest industrial sector in the park. Biotechnology, with only seven companies, is considered the new kid on the block, but the number of companies in this field is expected to expand significantly with the new phase of development.



Convenient access is a major attraction for Hsinchu residents.

The total number of employees in the park at the end of 1993 was 28,500: nearly 48 per cent have received at least a junior college or technical college education. Returning expatriates have played a vital role in the growth of the park.

By promotion through a network of overseas offices of the National Science Council and overseas Chinese profession-

al organizations, the park has gained recognition among the large population of Chinese overseas, particularly in the United States. By 1993 more than a thousand were working in the park, most of them from Silicon Valley companies, and 73 of the park companies had been founded by members of this group, or about half of the companies in the park. With the knowledge and ideas they have brought back, they have done much to raise the overall levels of technology and research and development capability in Hsinchu.

This overseas network of Chinese scientists and organizations is likely also to play an important role in supporting further expansion of the park.

The park's companies are expected to reach combined sales of US\$6 billion by 1996. In order to accommodate their rapid growth, the park is undertaking land development for the third and fourth phase expansions. New facilities are being planned by integrated circuit manufacturers, and the administration is working with the industrial and research sectors to bolster the development of the biotechnology industry. Furthermore, the government has approved the Hsinchu Science City project, with an investment of at least US\$4 billion over the next 25 years to develop Hsinchu into a true metropolitan science city.

Over the past 13 years, the efforts of many people have given the Hsinchu Science-based Industrial Park a solid foundation for success. Facing the challenges of stiffer environmental protection regulations, rising labour and operational costs, and the rise in value of the New Taiwan Dollar, Taiwan's industries must move towards more technology-intensive, capital-intensive, and high-value-added production. The science park's leading role in the changing industrial structure has made it possible for Hsinchu to become a major centre of high-tech industry in the Asia-Pacific region at the approach of the twenty-first century.

Steve Hsieh is Director General, Hsinchu Science Park Administration, 2, Hsin Ann Road, Hsinchu, Taiwan 30077.

INTERNATIONAL FACTS AND FIGURES

	North American parks	Non-North American parks	Worldwide total
Hectares	31,043	8,919	39,962
Hectares improved	15,849	3,491	19,340
Companies	4,746	6,369	11,115
Employees	255,425	178,112	433,537
Buildings	1,702	2,200	3,902
Buildings (m ² in parks)	10,932,414	5,555,932	16,488,346
Hectares of infrastructure	1,513	539	2,052
Hectares with structures	3,345	748	4,093
Hectares leased	1,170	326	1,496
Hectares available lease only	1,410	122	1,532
Hectares sold	4,664	477	5,141
Hectares available sale only	3,359	277	2,636
Hectares available sale or lease	2,234	306	2,540
Hectares under option	104	95	199

The data listed for hectares available for sale, hectares available for lease and hectares available for sale or lease reflect improved hectares.

- The worldwide count for science parks is 403. North American parks increased by 9 and Non-North American parks increased by 60 since 1991. The increase for Non-North American parks is attributed to better access to park information rather than newly established parks.
- North American parks surpassed 50 per cent of the world total for the first time, and have more improved land than unimproved land designated as a research park. Non-North American parks show a 39 per cent to 61 per cent comparison between improved and unimproved land.
- Since 1991, the number of companies in North American parks grew by 68% and the number of employees by 27%. Non-North American growth rates are not computed, due to incomplete data.
- The average size company in North American parks has approximately 54 employees. In Non-North American parks, the average size company is smaller, with 28 employees.

- Land dedicated to infrastructure makes up approximately 10 per cent of improved land in North American parks and over 15% in Non-North American parks. This seems to indicate that Non-North American parks allocate a greater amount of land per hectare for infrastructure. However, when taking into account the average size of improved hectares (14 hectares in Non-North American parks and 103 hectares in North American parks), North American parks provide a greater amount of total infrastructure to support their parks.

- The supply of available land either improved or unimproved in research and science parks is substantial worldwide.

These data were compiled by the International Association of Science Parks, based in the Bordeaux Technopolis Centre, and the Association of University Related Research Parks/4500 South Lakeshore Drive, Suite 475, Tempe, Arizona 85282.